VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

Fakulta elektrotechniky
a komunikačních technologií

DIPLOMOVÁ PRÁCE

Brno, 2021 Bc. Martin Šulič

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH TECHNOLOGIÍ
FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION

ÚSTAV TELEKOMUNIKACÍ
DEPARTMENT OF TELECOMMUNICATIONS

ZABEZPEČENÁ ARCHIVACE DAT S VYUŽITÍM CLOUDOVÉHO VÝPOČTU
SECURE DATA ARCHIVING USING CLOUD COMPUTING

DIPLOMOVÁ PRÁCE
MASTER'S THESIS

AUTOR PRÁCE — **Bc. Martin Šulič**
AUTHOR

VEDOUCÍ PRÁCE — **Ing. Vlastimil Člupek, Ph.D.**
SUPERVISOR

BRNO 2021

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH TECHNOLOGIÍ

Diplomová práce

magisterský navazující studijní program **Informační bezpečnost**

Ústav telekomunikací

Student: Bc. Martin Šulič ***ID:*** 195837
Ročník: 2 ***Akademický rok:*** 2020/21

NÁZEV TÉMATU:

Zabezpečená archivace dat s využitím cloudového výpočtu

POKYNY PRO VYPRACOVÁNÍ:

V diplomové práci analyzujte možnosti realizace privátního cloudu a zabezpečené dlouhodobé archivace dat pomocí open-source. Na základě provedených analýz navrhněte kryptograficky zabezpečený cloud zajišťující integritu, autentičnost, nepopiratelnost, časové ukotvení a důvěrnost archivovaných dat. Cloud bude vytvářet logy dokumentující činnosti uživatele (přihlášení/odhlášení uživatele, zápis/čtení dat apod.) a bude podporovat dlouhodobou interpretaci archivovaných dat. Implementujte Vámi navržené řešení, definujte jeho hardwarové nároky, ověřte jeho funkčnost, přehledně prezentujte jeho možnosti použití a ohodnoťte jeho kryptografickou bezpečnost.

DOPORUČENÁ LITERATURA:

[1] VARGHESE, Blesson; BUYYA, Rajkumar. Next generation cloud computing: New trends and research directions. Future Generation Computer Systems, 2018, 79: 849-861.

[2] HUTAŘ, Jan; MELICHAR, Marek. Dlouhodobá archivace digitálních dat–od teoretických úvah k praktické realizaci?. Knihovna, 2015, 26.2.

Termín zadání: 1.2.2021 ***Termín odevzdání:*** 24.5.2021

Vedoucí práce: Ing. Vlastimil Člupek, Ph.D.

doc. Ing. Jan Hajný, Ph.D.
předseda rady studijního programu

ABSTRAKT

Táto diplomová práca sa zaoberá podrobnou anlýzou možností realizácie privátneho cloudu a zabezpečenej archvácie dát na dlhé obdobie prostredníctvom nástrojov s otvoreným zdrojovým kódom. Popisuje jednotlivé štandardy a procesy prípravy dát, ako aj referenčný model OAIS pre dlhodobé uchovávanie. Z analyzovaných informácií je vytvorený kompletný návrh konečného riešenia, s popisom funkčnosti a spôsobom nasadenia v prostredí kontajnerov Docker. Implementácia návrhu a hlavná funkcionalita jednotlivých systémov ako Archivematica či Nextcloud je dôkladne popísaná a taktiež sú definované hardvérové nároky a ohodnotená kryptografická bezpečnosť.

KĽÚČOVÉ SLOVÁ

Archivácia, dlhodobé uchovávanie, cloud, kontajnerizácia, Docker, OAIS, Archivematica, Nextcloud

ABSTRACT

This master's thesis is focused on detailed analysis of possibilities of implementing a private cloud and secure data archiving for a long period of time using open-source tools. It describes the individual standards and processes of data preparation, as well as the OAIS reference model for long-term preservation. From the analyzed information, a complete design of the final solution is created, with a description of the functionality and the method of deployment in the environment of Docker containers. The design implementation and the main functionality of individual systems such as Archivematica or Nextcloud are thoroughly described and also the hardware requirements and cryptographic security were evaluated.

KEYWORDS

Archiving, long-term preservation, cloud, containerization, Docker, Archivematica, Nextcloud

ŠULIČ, Martin. *Zabezpečená archivace dat s využitím cloudového výpočtu*. Brno, 2021, 126 s. Diplomová práca. Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav telekomunikací. Vedúci práce: Ing. Vlastimil Člupek, Ph.D.

VYHLÁSENIE

Vyhlasujem, že svoju diplomovú prácu na tému „Zabezpečená archivace dat s využitím cloudového výpočtu“ som vypracoval samostatne pod vedením vedúceho diplomovej práce, s využitím odbornej literatúry a ďalších informačných zdrojov, ktoré sú všetky citované v práci a uvedené v zozname literatúry na konci práce.

Ako autor uvedenej diplomovej práce ďalej vyhlasujem, že v súvislosti s vytvorením tejto diplomovej práce som neporušil autorské práva tretích osôb, najmä som nezasiahol nedovoleným spôsobom do cudzích autorských práv osobnostných a/alebo majetkových a som si plne vedomý následkov porušenia ustanovenia § 11 a nasledujúcich autorského zákona Českej republiky č. 121/2000 Sb., o práve autorskom, o právach súvisiacich s právom autorským a o zmene niektorých zákonov (autorský zákon), v znení neskorších predpisov, vrátane možných trestnoprávnych dôsledkov vyplývajúcich z ustanovenia časti druhej, hlavy VI. diel 4 Trestného zákonníka Českej republiky č. 40/2009 Sb.

Brno ....

podpis autora

POĎAKOVANIE

Rád by som touto cestou poďakoval Ing. Vlastimilovi Člupkovi, Ph.D. za jeho odborné vedenie, konzultácie a trpezlivosť pri vedení mojej diplomovej práce. Taktiež chcem poďakovať svojim rodičom, bez ich podpory by táto práca nevznikla. V neposlednej rade by som chcel poďakovať všetkým priateľom a verným spolužiakom, kde vďaka patrí hlavne Bc. Samuelovi Sidorovi, Bc. Radoslavovi Heribanovi, Bc. Martinovi Hrdému, Bc. Ondřejovi Kupkovi a Bc. Tatiane Novosadovej, bez ktorých by som celé štúdium nedotiahol do konca.

Obsah

Zoznam obrázkov

Zoznam tabuliek

Úvod

Bezpečné uchovávanie informácií je dlhodobým problémom. V minulosti sa využívalo uloženie na tajné, alebo zabezpečené miesta (napríklad sejf). Inak tomu nie je ani v dnešnej dobe. Fyzickú držbu dát vystriedala digitálna doba, ale problém zostal rovnaký, dokonca sa ešte zhoršil. Celé svoje dianie ľudstvo zaznamenáva a presúva do digitálnej podoby, kedy sa neustále zväčšuje množstvo informácií, ktoré treba nielen uchovať, ale aj zabezpečiť. Tak isto ako kvantum digitálnych dát, sa zväčšuje aj počet hrozieb a okruh potencionálnych útočníkov, pretože sa už nemusí jednať len o subjekty, ktoré majú fyzický prístup k uloženým informáciám. Trendom v tejto dobe je presúvanie digitálnych informácií do ešte otvorenejšieho prostredia cloudov. Tým vzniká ešte väčší záujem o bezpečnosť v tejto oblasti.

Táto práca sa venuje bezpečnému uloženiu dát na privátnom cloude s podporou dlhodobej archivácie. Hlavnou témou je priblíženie techník dlhodobej archivácie podľa najnovších štandardov, ktoré sa využívajú vo veľkých repozitároch, a ich transformácia do podoby použitia lokálneho rázu. Práca podrobne popisuje možnosti digitálneho uchovávania dát, kde sa zameriava na jednotlivé štandardy, zvyklosti a možnosti celého procesu dlhodobej archivácie. Túto problematiku riešia už dlhé roky rôzne organizácie, spoločenstvá a spoločnosti, ktorých výsledkom sú referenčné modely, normy, ako aj rôzne odporúčania pri vytváraní digitálnych archívov pre dlhodobé uchovávanie dát.

Práca taktiež upozorňuje na potrebu zabezpečenia dát, ktorá začína už pri výbere typov jednotlivých fyzických médií, cez použité súborové systémy až po techniky pre zabezpečenie dôvernosti obsahu. Súčasťou je aj dôkladná analýza riešení pre privátne cloudy a archivačné systémy, ktorá využíva výhradne softvér s otvoreným zdrojovým kódom.

Na základe všetkých zistení bol vypracovaný návrh privátneho cloudu s podporou dlhodobej archivácie. Od neho sa odvíjajú jednotlivé testované koncepty použitia pre výber najlepšej varianty a typu archivačného systému. Implementácia konečného riešenia prebehla v prostredí kontajnerizácie s dôrazom na optimalizáciu hardvérových nárokov riešenia. Postup prípravy a nasadenia vybraného riešenia je dôkladne zdokumentovaný, kde základná funkcionalita jednotlivých komponentov je predstavená budúcemu používateľovi či správcovi. Súčasťou je taktiež zhodnotenie bezpečnosti dát a analýza hardvérových nárokov konečného riešenia.

Celkovo má práca dať čitateľovi do povedomia problematiku dlhodobej archivácie a taktiež možnosti, ako ju vykonávať efektívne, podľa štandardov, a za čo najmenšie vynaložené úsilie a finančné prostriedky.

1 Cloudový výpočet

Cloudový výpočet (Cloud Computing) označuje poskytovanie IT služieb, aplikácií, výpočtového výkonu ako aj úložiská dát na diaľku, v situácií, kedy používatelia nepotrebujú fyzický server alebo úložisko alokované u nich. Poskytovateľ cloudového výpočtu zvyčajne spravuje hardvér a softvér, a zároveň ponúka rôzne druhy zdieľaných služieb na vyžiadanie používateľom. Tým sa zvyšuje efektivita, kedy klient využíva len to čo potrebuje a po aký čas to potrebuje. Poskytované služby sa delia do troch hlavných kategórií, a to služby softvéru (SaaS), platformy (PaaS) a infraštruktúry (IaaS) [1].

Software as a Service (SaaS): Sa zameriava na koncových používateľov cloudu. Namiesto inštalovania aplikácií na zariadení klientov, SaaS aplikácie sú hosťované na cloudových serveroch a zároveň prístupné cez internet v podobe webového alebo programového užívateľského rozhrania. Táto kategória je najrozšírenejšia práve z dôvodu, že používateľ sa nestará o to, ako a na akej infraštruktúre daná aplikácia beží [2].

Platform as a Service (PaaS): Sa zameriava na potreby vývojárov aplikácií. Poskytovatelia týchto služieb ponúkajú všetko potrebné od vývojárskych nástrojov, operačných systémov až po samotnú infraštruktúru (výpočtový výkon) cez internet. Pre vývojárov je tento spôsob vývoja výhodný, pretože nepotrebujú si zaobstarávať potrebný softvér a hardvér, ale si ho iba prenájmu na presne potrebnú dobu. Poskytovateľ vie taktiež promptne prispôsobovať výpočtový výkon aktuálnym potrebám vývojára [2].

Infrastructure as a Service (IaaS): Sa zameriava na firmy alebo jednotlivcov, ktorí potrebujú priamy prístup k virtualizovanému alebo kontajnerizovanému hardvéru. Klient si na základe svojich požiadaviek (počet jadier procesora, veľkosť pamäte, atď.) prenajme od poskytovateľa potrebný výpočtový výkon vo forme virtuálneho stroja [2].

1.1 Verejný cloud

Verejný cloud (Public cloud) je sada hardvéru, sietí, úložiska, služieb, aplikácií a rozhraní vlastnená a prevádzkovaná komerčne treťou stranou, ktorej využitie poskytuje iným spoločnostiam alebo jednotlivcom cez internet. Poskytovatelia vytvárajú vysoko škálovateľné dátové centrá, ktorých základnú infraštruktúru používateľ nepozná. Vďaka dostupnosti veľkého množstva výpočtového výkonu, ktoré je hneď k dispozícií, si môžu zákazníci presne navrhnúť a optimalizovať zdroje potrebné pre beh ich aplikácie v tomto cloude, a tak znížiť ich potencionálne náklady. Väčšina

poskytovateľov ponúka aj širokú škálu rozhraní API (Application programming interface) a rôznych služieb, ktoré sú taktiež dostupné spôsobom na požiadanie.

V poslednej dobe sa rozširuje aj ponuka prenájmu konkrétneho hardvéru v datacentre. Zákazník tak má plnú kontrolu nad celým zariadením (nebude ho s nikým iným zdieľať), väčšinou serverom. Výhodou tohto riešenia je, že chod, správu a sieťovanie tohto zariadenia presúva zákazník na poskytovateľa verejného cloudu (datacentra). Nevýhodou sú nie až také veľké úspory prevádzkových nákladov v porovnaní s klasickým riešením [1], [3].

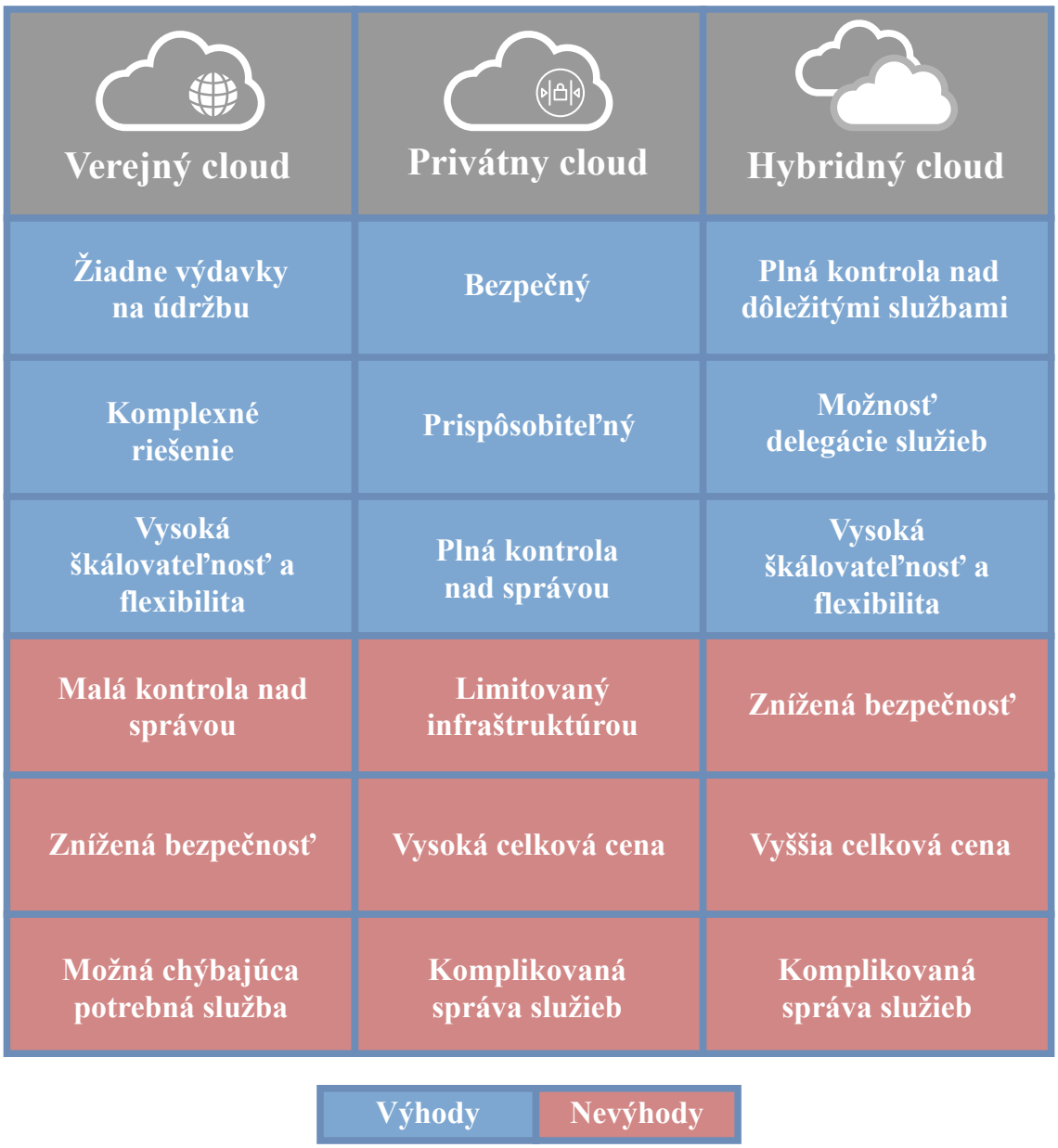
1.2 Privátny cloud

Privátny cloud (Private cloud) je sada hardvéru, sietí, úložiska, služieb, aplikácií a rozhraní vlastnená a prevádzkovaná organizáciou pre svojich zamestnancov, partnerov, alebo zákazníkov. Môže byť ale taktiež vytvorený a prevádzkovaný treťou stranou, ale výhradne iba pre jeden podnik. Privátny cloud je vysoko kontrolované prostredie skryté pred verejnosťou za firewallom.

Častým príkladom implementácie je, že poskytovateľ verejného cloudu zabalí svoje ponúkané služby do softvérového balíka, ktorý je následne nainštalovaný na hardvér konkrétnej organizácie. Tým dochádza k možnosti využitia skoro všetkých služieb verejného cloudu, ale lokálne v bezpečnom prostredí u zákazníka. Výhodou je, že organizácia má plnú kontrolu nad správou privátneho cloudu, ale zároveň ušetrí prostriedky na vývoj takéhoto riešenia [4], [5].

1.3 Hybridný cloud

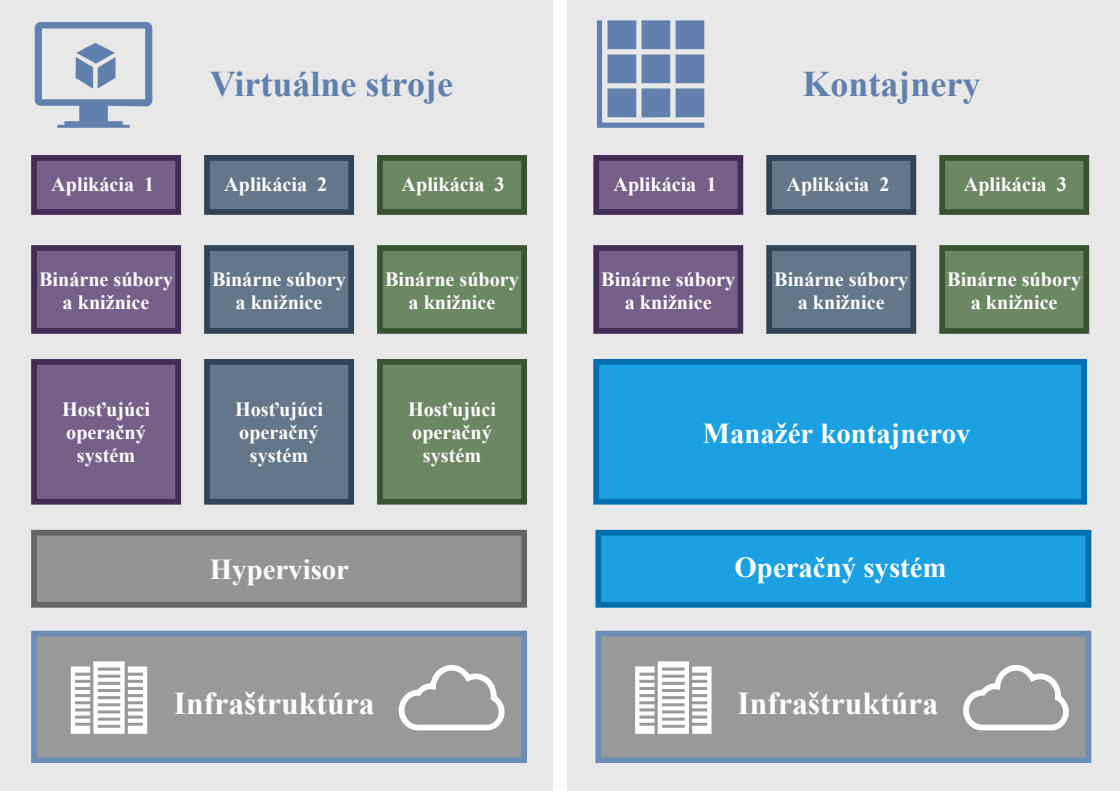
Hybridný cloud (Hybrid cloud) je kombináciou verejného a privátneho cloudu. Základom je privátny cloud spravovaný organizáciou lokálne za firewallom, ktorý má implementované niektoré služby z verejného cloudu. Podstatou je, že organizácie môžu niektoré špeciálne služby delegovať na poskytovateľov verejných cloudov, a tak ušetriť značné prostriedky. Výhodou je taktiež možnosť integrovania a prepojenia jednotlivých služieb od rôznych poskytovateľov s privátnym cloudom [5].



Obr. 1.1: Zhodnotenie parametrov cloudov.

2 Virtuálne prostredie

Virtuálne prostredie v oblasti serverov, vzniklo ako odpoveď na potrebu efektívnejšieho využívania výpočtového výkonu. Bežiace aplikácie a služby nedokážu vždy efektívne využiť všetky dostupné zdroje. Tento problém dokážu riešiť viaceré typy mechanizmov na hosťovanie aplikácií v počítačových systémoch, ako sú virtualizácia a kontajnerizácia, kde virtualizácia používa virtuálny stroj ako základnú jednotku, zatiaľ čo kontajnerizácia využíva koncept kontajnera [6], [7].



Obr. 2.1: Štruktúra virtualizačných systémov [7].

2.1 Virtualizácia

Virtualizácia pomáha vytvárať softvérovo založenú (virtuálnu) verziu počítačových zdrojov, ktoré zahŕňajú výpočtové jednotky, úložiská, siete, celé servery či dokonca aplikácie. Dovoľuje rozkúskovať jedno fyzické zariadenie do viacerých virtuálnych strojov. Každý z nich pracuje samostatne na vlastnom, často odlišnom operačnom systéme simultánne, pričom zdieľajú rovnaké zdroje.

K sprostredkovaniu virtualizácie je potrebný špecializovaný softvér alebo firmvér nazývaný hypervisor, ktorý má na starosti najdôležitejšiu časť, a to komunikáciu s

hardvérom a rozdelenie jeho prostriedkov medzi virtuálne stroje, viď obr. 2.1. Takže keď používateľ alebo program potrebuje dodatočný výpočtový výkon na virtuálnej stanici, požiadavka smeruje práve na hypervisor, ktorý si ju zapamätá, zhodnotí a posunie fyzickému systému [8].

2.2 Kontajnerizácia

Kontajnerizácia je odľahčená alternatíva k virtualizácií, ktorá zahŕňa zabalenie aplikácií do kontajnerov s jeho vlastným operačným prostredím, viď tab. 2.1. Takže namiesto inštalácie operačného systému pre každý virtuálny stroj zvlášť, kontajnery využívajú hostiteľský operačný systém. Každý kontajner je samostatný spustiteľný softvérový balík, ktorý beží nad hostiteľským systémom. Ten môže obsluhovať veľa kontajnerov súčasne, ktoré bežia ako izolovaný proces ku ktorému nemajú napriamo ostatní prístup. Toto zdieľanie systémových zdrojov, znižuje potrebu reprodukovať kód operačného systému, čo znamená, že na tom istom hardvéri je možné spustiť až trojnásobok aplikácií oproti virtuálnym strojom. Taktiež keďže kontajnery sú jednoduchšie, menšie (často megabajty), tak sú schopné rýchleho štartu, v rádoch niekoľkých sekúnd. Nevýhodou je, že daná aplikácia musí podporovať zabalenie do takéhoto kontajneru. Štruktúra takéhoto systému je oproti virtualizácií nasledovná. Hardvér neovláda hypervisor ale nahrádza ho jeden spoločný hostiteľský operačný systém. Nad ním beží manažér kontajnerov, ktorý sa stará o beh, alokáciu zdrojov a správu kontajnerov. V jednotlivých kontajneroch bežia samostatné aplikácie. Najznámejší manažér kontajnerov s otvoreným kódom je Docker. Vďaka jeho rozšírenosti a veľkého počtu podporovaných aplikácií je vynikajúcim nástrojom pre infraštruktúru kontajnerov [9].

Tab. 2.1: Porovnanie funkcionalít virtualizácií.

Virtuálny stroj	**Kontajner**
Plná virtualizácia	Ľahká virtualizácia
Zložitosť zväčšuje výkonnostné nároky	Jednoduchosť zmenšuje výkonnostné nároky
Na každom virtuálnom stroji beží samostatný operačný systém	Všetky kontajnery zdieľajú jeden rovnaký operačný systém
Virtualizácia na úrovni hardvéru	Virtualizácia na úrovni operačného systému
Štart za niekoľko minút	Štart za niekoľko sekúnd
Potreba väčšieho množstva pamäte	Vyžaduje menej pamäte
Plná izolácia	Izolácia na úrovni procesov

2.3 Docker

Docker je platforma alebo skôr technológia s otvoreným zdrojovým kódom pre kontajnerizáciu. Umožňuje spustenie aplikácie v izolovanom prostredí, ktoré je sformované do kontajneru. Na jednom hostiteľovi je súčasne možné spustenie viacerých na seba nezávislých alebo závislých kontajnerov, ktoré obsahujú všetko potrebné na spustenie danej aplikácie. Odpadá tak starosť o to, čo je nainštalované na hostiteľovi. Tým sa zabezpečuje flexibilita pri vytváraní, nasadzovaní, kopírovaní či presúvaní kontajnerov z prostredia do iného prostredia, čo vedie k optimalizovaní aplikácií pre cloud.

Počiatočné verzie Dockeru využívali LXC (Linux Containers) kontajnery, ale neskôr bola vyvinutá vlastná technológia, ktorá sa odlišuje hlavne v tom, že docker kontajnery dokážu bežať bez potrebných modifikácií na rôznych platformách či zariadeniach. Ďalej oproti LXC dokáže byť kontajner automaticky zostavený priamo zo zdrojového kódu. Výhodou je tiež, zostavenie aplikácie z viacerých častí, kedy každá z nich je spustená vo vlastnom docker kontajneri, a tým je umožnená čiastková oprava alebo aktualizácia bez nutnosti vypínania celej aplikácie [10].

2.3.1 Dockerfile

Je jednoduchý textový súbor ktorý obsahuje inštrukcie respektíve zoznam príkazov pre Docker Engine ako vytvoriť obraz (image) pre budúci kontajner. Dockerfile automatizuje tento proces tvorby. Tento súbor nemá žiadnu príponu [11].

2.3.2 Docker image

Tento obraz obsahuje spustiteľný zdrojový kód aplikácie so všetkými potrebnými nástrojmi, knižnicami. Vytvára sa ako predloha budúceho kontajneru buď priamo zo zdrojového kódu, alebo sa sťahujú predpripravené obrazy zo spoločných úložísk (napríklad Docker Hub). Vždy keď sa zmení niečo v obraze, jeho najnovšia verzia je zvyčajne dostupná pod tagom latest. Tag oddeľuje dvojbodka za názvom obrazu, a využíva sa pre odlíšenie starších verzií, ktoré sa stále môžu využívať (presne ako napríklad staršie verzie programu) [11].

2.3.3 Docker container

Kontajnery sú bežiace celky podľa obrazov, v ktorých je spustená aplikácia alebo jej časť. Zatiaľ čo Docker obrazy sú len súbory na čítanie a predstavujú predlohu, s kontajnermi môžu správcovia interagovať (napríklad cez konzolu), upravovať ich nastavenia a podobne. Docker vytvára kontajnery na základe vlastného štandardu,

takže dokážu byť veľmi ľahko prenosné a ich spustenie záleží len od toho, aby na danom zariadení bežal Docker Engine, viď obr. 2.2. Z existujúceho kontajneru sa môže tak isto vytvoriť predloha (obraz) pre ďalšie nové kontajnery [11].

2.3.4 Docker volume

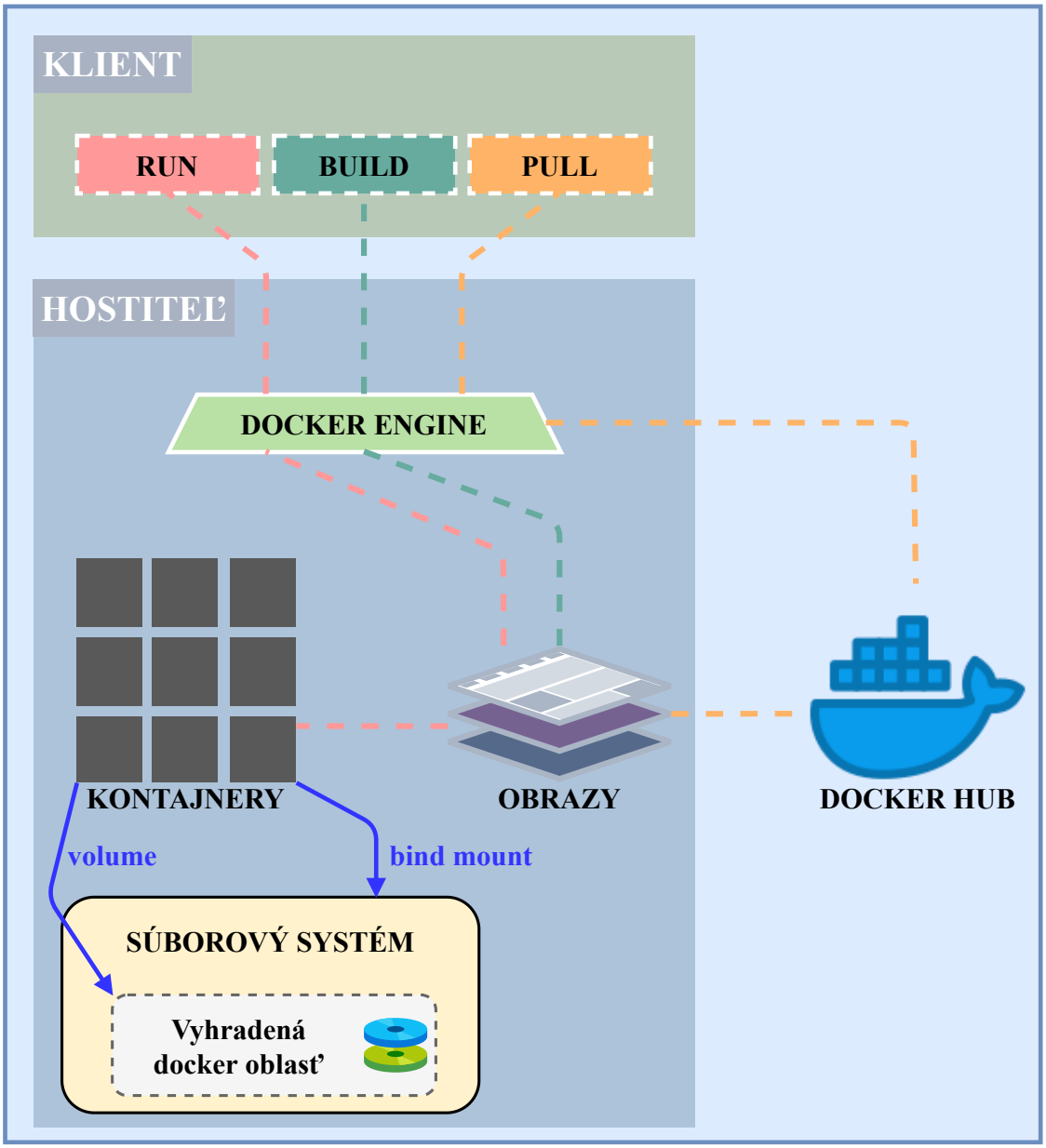
Keď sa z obrazu vytvorí kontajner, všetky následné zmeny v ňom, napríklad pridanie alebo odstránenie súborov, majú predvolenú životnosť len počas behu daného kontajneru. Keď je kontajner poškodený, zničený alebo vymazaný, odstraňujú sa s ním aj všetky dáta. Na uchovávanie údajov, ktoré je potrebné mať k dispozícií nezávislé na behu kontajnera, sa používajú techniky mapovania zväzkov (Docker volume), viď obr. 2.2. Tie taktiež fungujú multiplatformovo, a ich obsah je možné zdieľať medzi viacerými kontajnermi. Ovládače týchto zväzkov ich dokážu uložiť na vzdialených hostiteľoch alebo cloudoch, šifrovať ich obsah a iné. Ak sa vyžaduje ukladať údaje na konkrétnom mieste na hostiteľovi alebo použiť už existujúce údaje na disku, dajú sa tieto umiestnenia taktiež pripojiť (Bind Mounting) [11].

2.3.5 Docker networking

Docker automaticky pri inštalácií vytvára tri základné siete: Bridge, Host, None. Primárnou sieťou pre všetky novovytvorené kontajnery je Bridge, ktorá prideľuje adresy v IP rozsahu 172.17.x.x, a tým sa stará o segregáciu od hostiteľskej siete. Pre prístup ku kontajnerom zvonka musia byť porty namapované na porty hostiteľa. Sieť Host neposkytuje žiadnu izoláciu a tak kontajner bude bežať na rovnakom porte hostiteľa. Nevýhodou je, že sa tak znemožňuje použitie rovnakého portu hostiteľa pre viacero kontajnerov. None sieť udržuje kontajner v celkovej izolácií, takže nie je pripojený k žiadnej sieti alebo kontajneru [11].

2.3.6 Docker hub

Je najväčšie verejné úložisko pre Docker obrazy. Všetci používatelia môžu zdieľať svoje obrazy ako ich aj sťahovať pre vytvorenie kontajnera. Výhodou je, že veľa firiem a developerov vytvára aktualizované obrazy svojich aplikácií, takže ich netreba tvoriť lokálne zo zdrojového kódu, ale si ich len stiahnuť ako predlohu. Treba ale upozorniť, že je dôležité používať prednostne obrazy od vývojárov, alebo používateľov, ktorí uvádzajú zdroj ako a z čoho boli zostavené. Neznáme obrazy predstavujú možné bezpečnostné riziko [10].



Obr. 2.2: Architektúra docker.

2.3.7 Docker compose

Pri zostavovaní komplexnejších riešení alebo aplikácií, kde sa využíva viacero na seba nadväzujúcich kontajnerov na jednom hostiteľovi, je jednoduchšie pre správu a zavedenie použiť nástroj Docker compose. Ten použije vytvorený YAML súbor (docker-compose.yml), v ktorom sú konfigurácie všetkých služieb (kontajnerov) aplikácie. Dajú sa v ňom definovať aj zväzky (volumes), sieťové prepojenie, závislosti na jednotlivých službách a mnoho ďalšieho. Docker compose teda umožňuje pomocou jedného príkazu vytvorenie a spustenie všetkých služieb z konfigurácie naraz [11].

2.3.8 Manažér kontajnerov

Práca s dockerom vyžaduje prevažne použitie konzole a textového editora. Pre neskúseného používateľa (správcu) je takáto práca zdĺhavá a neprehľadná. Preto existujú nástroje na manažovanie kontajnerových aplikácií, ktoré obsahujú grafické rozhranie v rôznych formách pre vizualizáciu docker kontajnerov a ich správu. Jedným z týchto nástrojov s otvoreným zdrojovým kódom je Portainer. Poskytuje používateľské rozhranie formou webu, a celý nástroj beží ako samostatný kontajner na Docker engine [12].

3 Digitálne uchovávanie dát

Digitálne uchovávanie spočíva v zabezpečení prístupu k digitálnemu materiálu v budúcnosti. Problémom je, že väčšina médií na uchovávanie má obmedzenú životnosť a hardvér, softvér ako aj formáty súborov môžu po čase zastarať. Túto problematiku riešia už dlhé roky rôzne organizácie, spoločenstvá, spoločnosti, atď., ktorých snahou je stanoviť dlhodobý štandard v tomto odvetví. Výsledkom ich práce sú referenčné modely, normy ako aj rôzne odporúčania pri vytváraní digitálnych archívov.

3.1 Zálohovanie a archivácia dát

Obidva pojmy majú odlišnú funkciu, ale v dnešnej dobe sa ich význam často zamieňa, alebo sú nepresne definované ich rozdiely a využitie. V nasledujúcej časti je použitie jednotlivých termínov vysvetlené a následne porovnané.

3.1.1 Zálohovanie

Záloha je kópia dát vytvorená pre ochranu proti ich strate. V prípade poškodenia primárnych dát sa požadované informácie obnovia zo zálohy. Zálohovať by sa mali všetky dáta, ktorých poškodenie by malo za následok nejakú ujmu. Preto sa väčšinou vykonáva na informáciách, s ktorými sa aktuálne pracuje, to znamená že uchovanie jednej zálohy nemá byť po dlhý čas, slúži teda len ako krátkodobý zdroj pre ľahkú obnovu po nečakanej situácií. Typicky sa zálohy vytvárajú v presných časových intervaloch, alebo vtedy, keď sa dáta zmenia. Originály zostávajú zachované, ale staršie zálohy, ktoré nie sú potrebné a majú svoju aktuálnejšiu verziu sa vymazávajú. Vytvorené kópie by sa mali ukladať najlepšie na úplne oddelené médium od toho prostredia, kde bol vytvorený originál, pretože v prípade kompromitácie celého systému, sú lokálne zálohy neúčinné.

Situácie kedy a prečo zálohovať sa pre priblíženie môžu rozdeliť napríklad do nasledujúcich skupín:

- ľudský faktor,
- zlyhanie hardvéru,
- cielený útok.

To, kedy zlyhá ľudský faktor sa nedá predikovať. Preto by sa mala vytvárať záloha vždy, kedy sa mení alebo pristupuje k obsahu dôležitých dát. Príkladom je rozpracovaný dokument človekom, ktorý zmení nejaký obsah, ale po uložení zistí, že sa pomýlil a informáciu ktorú zmenil chce vrátiť späť. Bez dostupnej zálohy by sa nemohol vrátiť do stavu pred zmenou. Ďalej to môže byť nechcené vymazanie časti dát, alebo zmena nastavení, ktorá k tomu dospeje.

Zlyhanie hardvéru je tiež nečakané, ale dá sa už z časti predikovať. Keď sa používa zariadenie, ktoré je na hranici životnosti, dá sa predpokladať jeho zlyhanie a preto aj z toho vyplývajúca dedukcia kompletnej zálohy potrebných dát. Typickým príkladom je zlyhanie pevného disku, celého zariadenia alebo aj výpadok elektrického prúdu.

Chyba systému sa dá očakávať hlavne pri aktualizáciách a vykonávaní všetkých úkonov, ktoré môžu spôsobiť jeho kolaps. Nečakané chyby s dopadom na stratu dát sú menej časté, ale stále reálne. Preto je odporúčané vytvorenie bodu obnovy pred každým zásahom do systému.

Cielený útok môžeme rozdeliť do dvoch kategórií. Prvou je využitie chýb systémov, programov alebo algoritmov útočníkom s možným následkom straty dát. Druhou je využitie nedbalosti administrátorov, správcov systémov alebo používateľov, ktorí vedome či nevedome vytvorili bezpečnostné riziko. Zvyčajne sa tento typ situácie spája so zlyhaním ľudského faktora. Príkladom môže byť ransomware útok, ktorý šifruje dáta a následne útočník vyžaduje zaplatenie za ich znovusprístupnenie [13].

3.1.2 Archivácia

Archív je kópia dát vytvorená pre ich dlhodobé uchovanie. Originálne dáta môžu, ale nemusia ostať zachované v zdroji po tom, čo sa vytvorí a uloží ich archívna kópia. Je úplne bežné, že sa dáta po archivovaní zo zdroja vymažú. Archivovať by sa mali všetky údaje, ktoré bude možno potrebné na niečo využiť v budúcnosti, ale aktuálne sa s nimi nebude aktívne narábať. Týmto procesom sa teda takpovediac odložia dáta, ktoré treba uchovávať z nejakého dôvodu dlhšiu dobu a uvoľní sa tak miesto na rýchlejších, viac využívaných súborových úložiskách, typicky účtovné informácie, zmluvy, ale aj výskumné výsledky, archívy webov a podobne. Vytvorený archív by mal byť v takom formáte, aby bol čitateľný aj po dlhšej dobe. Túto vlastnosť dokážu zabezpečiť archivačné systémy, ktoré spracovávajú údaje podľa medzinárodných štandardov. Častou požiadavkou je aj dokázanie autentičnosti archivovaných dát [13], [14].

3.1.3 Porovnanie a simultánne využitie

Aj keď sú tieto dva prístupy ku kópiám dát odlišné, najlepšie je ich využívať simultánne. Pre dosiahnutie najlepšej bezpečnosti a zároveň efektivity uloženia dát je potreba dobre nastaviť proces zálohovania, kedy sa zálohujú všetky dôležité údaje ako ochrana pred ich náhlou stratou. Tento proces dopĺňa archivácia, vďaka ktorej sa zbytočne nezahlcuje systém nepoužívanými ale potrebnými dátami a tým znižuje

Tab. 3.1: Archivácia vs. zálohovanie.

Záloha	Archív
Umožňuje obnovu aktívne používaných dát	Uchováva dáta ktoré sa už aktívne nevyužívajú ale musia byť k dispozícií
Veľká rýchlosť prístupu k zálohe a následnej obnove	Veľká rýchlosť prístupu k celému archívu nie je dôležitá
Jedna z viacerých kópií údajov, môže poskytovať viac verzií (obnovenie na základe výberu času verzie)	Zvyčajne posledná kópia dát
Nove kópie zvyknú prepisovať staré, neaktuálne verzie zálohy	Údaje nie je možné len tak meniť či mazať
Krátkodobé uchovanie dát, uchovávajú sa tak dlho pokým sa aktívne využívajú	Dlhodobé uchovanie dát na určené obdobie alebo na neurčito

riziko ich straty, viď tab. 3.1. Oba varianty uchovávania dát sú teda spoločne dôležité pre bezpečnú stratégiu správy dát [13].

Keďže sa môže stať, že zlyhajú umiestnenia so zálohami a archívmi, ako doplnok bezpečnosti sa využíva takzvaná technika zotavenia po havárií (Disaster Recovery), čo je vlastne funkcia, ktorá duplikuje (automaticky zrkadlí) dáta na iné nezávislé úložisko. Tento krok je dôležitý hlavne pri archívoch, pretože tie na rozdiel od záloh nemajú už inú kópiu dát, a pri strate archívu by nastala celková strata všetkých dát. Často sa volí ako cieľové úložisko verejný cloud, ktorý dokáže zabezpečiť viacero odlišných geografických lokácií pre tieto duplikáty a tým aj rapídne zvýšiť bezpečnosť proti náhodnej strate [15].

3.2 Úrovne digitálneho uchovávania podľa NDSA

Konzorcium NDSA (National Digital Stewardship Alliance) združujúce viac ako 250 organizácií za účelom spolupráce pri projektoch dlhodobého uchovávania digitálnych informácií vydalo v roku 2018 dokument Levels of Digital Preservation (LoP - Úrovne digitálneho uchovávania), ktorý má za úlohu sprostredkovať základné rady organizáciám pri vytváraní a manažovaní svojich digitálnych archívov. Delí sa na štyri základné úrovne, ktoré popisujú funkčné oblasti úložiska dát, integrity dát, metadát, kontroly a obsahu dát [16].

Tab. 3.2: Úrovne digitálneho uchovávania podľa NDSA.

	1 – Poznajte svoj obsah	**2 – Chráňte svoj obsah**	**3 – Sledujte svoj obsah**	**4 – Udržujte svoj obsah**
Úložisko dát	Existujú dve úplne kópie dát uložené v stabilných úložiskách na separátnych lokalitách. Dokumentujú sa všetky médiá, kde je obsah uložený.	Uložené sú tri úplné kópie dát, s najmenej jednou v samostatnom geografickom umiestnení. Dokumentujú sa úložiská dát a média vrátane potrebných informácií k ich použitiu.	Aspoň jedna kópia dát je uložená v samostatnom geografickom umiestnení s rozdielnou hrozbou katastrofy ako ostatné kópie. Aspoň jedna kópia je uložená na rozdielnom type média. Sleduje sa zastarávanie úložiska a médií.	Uložené sú tri úplné kópie dát v samostatných geografických umiestneniach s rozdielnou hrozbou katastrofy. Maximalizuje sa diverzifikáciu úložiska, aby ste sa vyhlo bodovým zlyhaniam. Je vypracovaný podrobný plán na riešenie zastarávania hardvéru a softvéru úložiska.
Kontrola	Sú vymedzené presné práva, kto a ako môže pristupovať a narábať s obsahom.	Prístupové oprávnenia k obsahu sú zdokumentované a aplikované.	Udržujú sa záznamy (logy) o tom, kto a ako narábal s obsahom.	Vykonávajú sa pravidelné kontroly logov o prístupe/akciách.

Integrita dát	Overenie integrity v prípade, že boli dáta dodané spolu s kontrolným súčtom. Ak kontrolné súčty nie sú súčasťou obsahu, sú pri prevzatí dát vygenerované. Prebieha vírusová kontrola obsahu, problémové dáta sa izolujú do karantény.	Pri premiestňovaní alebo kopírovaní obsahu sa overuje jeho integrita. Originálne médiá sú blokované proti zápisu. Informácie o integrite sú zálohované na inom mieste ako obsah.	V stanovených intervaloch sú verifikované informácie o integrite. Udržujú sa záznamy (logy) o stave integrity dát, na požiadanie je možné dodať audit týchto dát.	Overujú sa informácie o integrite v reakcii na konkrétne udalosti alebo činnosti. Podľa potreby sa vymieňa alebo opravuje poškodený obsah.
Metadáta	Je vytvorený inventár obsahu a tiež sa dokumentuje súčasná lokácia úložiska. Inventár je zálohovaný s aspoň jednou kópiou uloženou oddelene od obsahu.	Je uložených dostatok metadát na to, aby sa vedel identifikovať obsah (môže to zahŕňať kombináciu administratívnych, technických, popisných, archivačných a iných metadát).	Sú doplnené potrebné metadáta aby plnili ustanovené štandardy.	Zaznamenáva sa čas a akcie spojené s archiváciou obsahu. Sú implementované vybrané štandardy metadát.

Obsah	Dokumentujú sa formáty súborov a ďalšie základné charakteristiky obsahu vrátane toho, ako a kedy boli identifikované.	Overujú sa formáty súborov a ďalšie základné charakteristiky obsahu. Sú budované vzťahy s tvorcami obsahu pre ustanovenie udržateľného formátu súborov.	Monitoruje sa zastaralosť a zmeny technológií, od ktorých závisí čitateľnosť obsahu.	Vykonávajú sa migrácie, normalizácie, emulácie a podobné činnosti, ktoré zabezpečujú prístup k obsahu.

3.3 OAIS

Referenčný model OAIS (Open Archival Information System) je štandard, ktorý vyvinutý výborom CCSDS (Consultative Committee for Space Data Systems) a označuje sa pod normou ISO 14721. Jeho obsah je komplexnejší oproti modelu NDSA, a preto je aj viac rozšírený. Je to vlastne koncepčný rámec s vysokou úrovňou abstrakcie a flexibility. Zavádza terminológiu a špecifikujte základné požiadavky, subjekty, vzťahy a procesy archivačného systému, ale nekladie konkrétne požiadavky na samotnú implementáciu ani použité technológie. Preto je aplikovateľný na akýkoľvek digitálny archív, ale hlavne pre tie, ktoré sa špecializujú na dlhodobé uchovávanie dát. OAIS model nie je len čistou predlohou pre archívy vyhovujúce tomuto štandardu, ale aj východiskovým bodom pre vývoj nových metodík, štandardov, koncepcií, terminológií a formátov vychádzajúcich práve z požiadaviek špecifikovaných v tomto modeli [17], [18].

3.3.1 Prostredie OAIS

Existujú tri základné role s určitými úlohami, ktoré predstavujú organizaciu či jednotlivca. V prostredí OAIS sú to tvorca, koncový užívateľ a manažment. Tvorca a koncový užívateľ môžu byť zastúpený pomocou iného pridruženého systému [17].

Tvorca

Posiela dáta do systému pre uchovanie. Ako prvé je jeho úlohou ustanoviť dohodu o podaní (Submission Agreement) s OAIS. Tá môže obsahovať špecifikáciu poskytnu-

tých dát, proces extrakcie metadát, ako aj autentifikačné údaje. Po jej ustanovení, posiela tvorca dáta do systému špeciálnym dátovým spojením (Data Submission Session) [18].

Koncový užívateľ

Chce dáta zo systému. Najskôr musí ustanoviť dohodu o objednaní dát (Order Agreement), ktorá môže byť v vybavená hneď alebo v periodických intervaloch. Tá identifikuje, ktorý balík informácií užívateľ požaduje, ako má byť transformovaný a následne odoslaný cez špeciálne dátové spojenie. Môže tiež obsahovať, kto presne dané dáta vyžaduje, jeho práva k nim, alebo iné potrebné informácie. Ak koncový užívateľ nemá presnú znalosť čo má požadovať od systému, ešte pred ustanovením dohody nadviaže spojenie pre vyhľadávanie (Search Session) s OAIS. Proces vyhľadávania býva iteračný, čo znamená, že široké informácie o dátach sa postupne spresňujú až kým nepríde k dostatočnej identifikácií záujmového balíku dát [18].

Manažment

Spravuje archívne politiky a definuje rozsah archívu. Spravidla vykonáva všetky potrebné riadiace činnosti (analýza rizík, riešenie konfliktov, atď.) [18].

3.3.2 Informačné balíky

Schopnosť transformovať dáta do takého tvaru, aby mohli byť archivované po dlhú dobu, je dosiahnutá pridaním dodatočných reprezentačných informácií (napríklad formát, štruktúra), ktoré sú s nimi uložené (Representation Information). Toto zoskupenie je definované ako informácie o obsahu (Content Information), ktoré spolu s informáciami o uchovaní (Preservation Description Information – PDI) a informáciami o zabalení (Packaging Information) vytvárajú balík informácií. PDI pozostávajú z informácií potrebných pre dosiahnutie požiadaviek archivácie. Príkladom je kontrolný súčet, ktorý sa vytvára pre potreby overenia integrity dát alebo povolenia k prístupu. Informácie o zabalení zase obsahujú poznatky, ktoré poukazujú na zviazanie PDI s informáciami o obsahu, uloženými na spoločnom úložisku [18].

Submission Information Package

Submission Information Package (SIP) je balík informácií zaslaný tvorcom do systému na archiváciu. Jeho forma a obsah je zvyčajne popísaná v dohode o podaní. Väčšina SIP obsahuje PDI a informácie o obsahu. Tvorca môže taktiež priložiť popisné informácie (Descriptive Information), ktoré sú späté s dátami a používajú sa na vyhľadávanie potrebného obsahu archivačných balíkov koncovým používateľom [18].

Archival Information Package

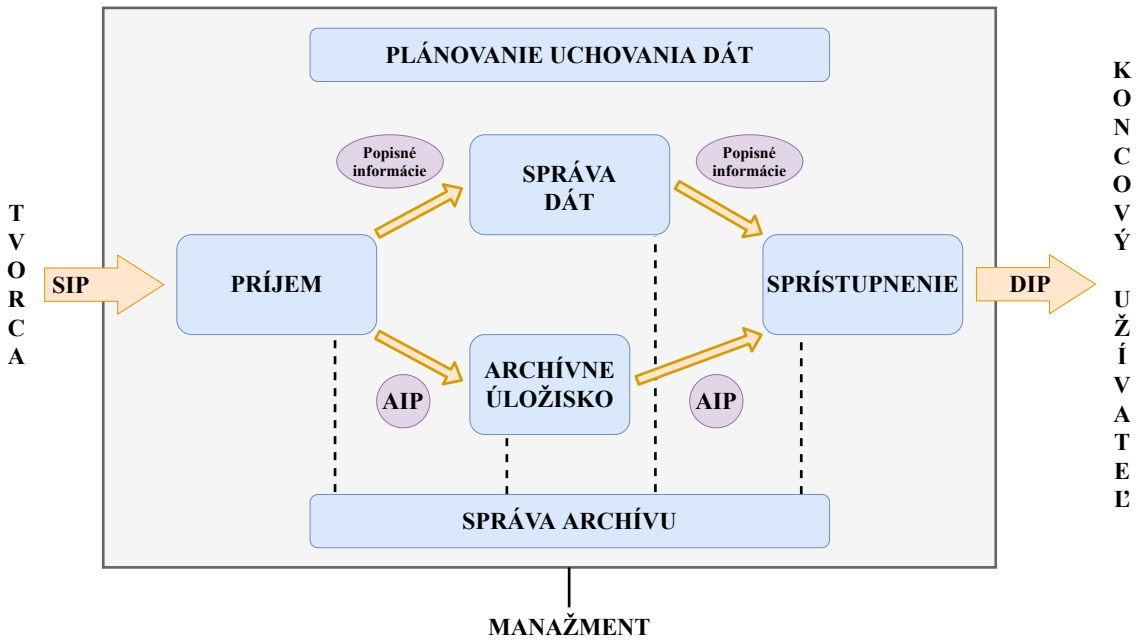
Archival Information Package (AIP) je verzia informačného balíka ktorá je uložená (archivovaná), a považuje sa za najdôležitejšiu časť. Vytvára sa pomocou transformácie balíka SIP v OAIS a obsahuje informácie o obsahu (dáta a ich reprezentačné informácie) a príslušné PDI. Z jedného SIP balíka sa môže transformáciou vytvoriť jeden ale aj viac AIP [18].

Dissemination Information Package

Dissemination Information Package (DIP) je balík, ktorý sa vytvára pri procese, kedy koncový užívateľ požiada o archivované dáta zo systému. Na základe dohody o objednaní dát sa jeden alebo viac určených AIP pretransformuje na DIP a následne sa zasiela žiadateľovi. Priložené PDI môžu, ale nemusia byť kompletné pri výstupe. informácie o zabalení musia byť prezentované koncovému užívateľovi v takej forme, aby vedel jasne odlíšiť jednotlivé dáta, ktoré požadoval [18].

3.3.3 Model funkčnosti (Funkčné entity)

Základom OAIS modelu funkčnosti je popis interakcie jednotlivých funkčných entít medzi sebou. Nasledujúci diagram 3.1 znázorňuje cestu a transformáciu balíkov informácií (Information Package – IP) jednotlivými entitami (procesmi, modulami) [17].



Obr. 3.1: OAIS model funkčnosti [17].

Príjem

Príjem (Ingest) je funkčná entita, ktorá prijíma balíky SIP od tvorcu a následne ich transformuje do podoby AIP. Môže sa tu vykonávať veľa procesov ako sú napríklad detekcie formátov, kontrolné súčty, vírusové skeny ako aj extrakcia metadát. Po vykonaní všetkých potrebných akcií je výsledný AIP uchovaný v archívnom úložisku a metadáta v module správy dát [17].

Archívne úložisko

Archívne úložisko (Archival Storage) je funkčná entita, ktorá poskytuje služby a funkcie pre úložisko a správu balíkov dát AIP. Po obdržaní týchto balíkov po procese transformácie ich uloží na permanentné úložisko a zistí kontrolu chýb, aktualizáciu hierarchie uloženia a dostupnosť pre proces (entitu) sprístupnenia [17].

Správa dát

Správa dát (Data management) obsahuje všetky popisné informácie o uchovávaných AIP a taktiež administratívne dáta potrebné pre správu celého archívu. Potrebné metadáta o AIP uchováva vo vhodnom tvare pre potrebu vyhľadávania obsahu koncovým užívateľom pomocou entity sprístupnenia [17].

Sprístupnenie

Sprístupnenie (Access) zabezpečuje komunikáciu s entitou správy dát a transformáciu archivovaných AIP na základe požiadavky od koncového používateľa. Stará sa teda o vyhľadávanie potrebných metadát a sprostredkováva ustanovenie dohody o objednaní dát, ktorej súčasťou sú aj informácie o oprávnení k prístupu do systému a jednotlivým AIP. Nasleduje transformácia týchto dát do formy balíka DIP a jeho doručenie koncovému užívateľovi [17].

Plánovanie uchovania dát

Cieľom plánovania uchovania dát (Preservation Planning) je zabezpečiť, aby archivované informácie zostali v tvare, ktorý bude zrozumiteľne čitateľný aj po dlhom čase. Táto entita teda zabezpečuje pravidelné vyhodnotenie aktuálnosti archivovaných informácií a celého systému k novým štandardom či technológiám a odporúča prípadné aktualizácie či migrácie obsahu do novej podoby. Zabraňuje tak informačnému zastarávaniu vďaka sledovaniu a analýze technologickej evolúcie a rizík s tým spätým. Preto sa predpokladá, že z dlhodobého hľadiska je táto funkcia zastávaná skôr prácou ľudí (komunitou) pomocou neustáleho výskumu ako automatizovaným procesom [17].

Správa archívu

Správa archívu (Administration) má na starosti komunikáciu medzi všetkými internými ako aj externými entitami. Je zodpovedná za monitorovanie systému, konfiguráciu a celkovú generálnu koordináciu vo vnútri systému. Taktiež dozerá na stanovenie a udržiavanie štandardov a politík archívu, ako aj poskytovanie používateľskej podpory. Príkladom je monitorovanie archívneho úložiska, ktorého výsledky správa archívu interpretuje správe dát, a tá ich môže predložiť koncovému užívateľovi cez entitu sprístupnenia, za predpokladu, že je prístup špecifikovaný v politikách poskytnutých manažmentom [17].

3.3.4 Predbežný príjem

Predbežný príjem (pre-ingest) je proces, ktorý nespadá priamo pod model funkčnosti OAIS, ale sa deje priamo pred ním. Týmto procesom tvorca vytvára z bežných dát balíčky SIP, tým že k nim prikladá príslušné metadáta a formuje ich na základe preddefinovaného štandardu. Takže vlastne neštruktúrované informácie sa stanú štruktúrovanými. Táto funkcia môže, ale nemusí byť súčasťou archivačného systému. Ak sa jedná o samostatné aplikácie ide o nástroje predbežného príjmu (pre-ingest tools) [19].

3.4 Metadáta

Metadáta, doslova „údaje o údajoch“, sú dnes široko používaným, ale často bližšie nešpecifikovaným pojmom, ktorý je chápaný rôzne medzi odbornými komunitami, ktoré navrhujú, vytvárajú, popisujú, chránia a používajú informačné systémy a zdroje [20]. Metadáta sú štruktúrované informácie, ktoré popisujú, vysvetľujú, lokalizujú, alebo inak uľahčujú získanie, použitie, alebo spravovanie zdroja informácií, viď tab. 3.3. Vytvárajú sa pre uľahčenie zisťovania relevantných informácií, môžu pomôcť aj pri organizovaní elektronických zdrojov, zlepšení interoperability a integrácie starších zdrojov, či poskytnutie digitálnej identifikácie ale aj podpore archivácie dát [21].

Podľa organizácie NISO (National Information Standards Organization) sa rozlišujú štyri základné druhy metadát [22]:

- Popisné metadáta – používajú sa pre vyhľadávanie a identifikáciu zdroja.
- Administratívne metadáta – delia sa na tri podkategórie.
  - Technické metadáta – použitie hlavne pri dekódovaní a vykresľovaní súborov.
  - Metadáta uchovania – podpora dlhodobého uchovávania súborov.

  - Právne metadáta – vyjadrenie práv duševného vlastníctva spojených s obsahom.
- Štrukturálne metadáta – zabezpečujú popis vzťahov častí zdrojov s inými.
- Značkovacie jazyky – integrujú metadáta a značky pre ďalšie štrukturálne alebo sémantické prvky v obsahu. Tieto jazyky väčšinou mixujú spolu dáta a metadáta ale len v spojení s ďalšími metadátami.

Tab. 3.3: Príklady a využitia obsahu metadát [21].

Typ metadát	**Vzorový obsah**	**Primárne využitie**
Popisné metadáta	Názov, Autor, Predmet, Žáner, Dátum publikácie	Objavenie, Zobrazenie, Interoperabilita
Technické metadáta	Typ súboru, Veľkosť súboru, Čas/Dátum vytvorenia, Komprimačná schéma	Interoperabilita, Uchovávanie, Manažment digitálneho objektu
Metadáta uchovania	Kontrolný súčet, Udalosť uchovávania	Interoperabilita, Uchovávanie, Manažment digitálneho objektu
Právne metadáta	Autorské práva, Licenčné podmienky, Držiteľ práv	Interoperabilita, Manažment digitálneho objektu
Štruktu-rálne metadáta	Sekvencia, Miesto v hierarchii	Navigácia
Značkovacie jazyky	Paragraf, Nadpis, Meno, Dátum	Interoperabilita, Navigácia

3.4.1 PREMIS

Preservation Metadata: Implementation Strategies (PREMIS) je jeden z hlavných štandardov v oblasti metadát pre podporu uchovávania a zabezpečenia dlhodobej čitateľnosti dát. Vyvinutý bol medzinárodným tímom odborníkov, preto je aj široko uznávaný a implementovaný v množstve nástrojov ako aj systémov pre digitálne uchovávanie. Zámerom je popísať vlastnosti digitálneho obsahu potrebného na podporu procesu archivácie (uchovávania), sledovať prijaté opatrenia a zaznamenávať informácie o zodpovedných aktéroch. PREMIS je teda schopný vyjadriť technické informácie o objektoch, ako aj ich históriu spracovania a správy [23], [24].

3.4.2 Dublin core

Dublin core (DC) je štandard pre metadáta, ktorý má za úlohu uľahčiť vyhľadávanie v elektronických dokumentoch. Je to schéma na ukladanie popisných informácií dátových objektov, ktorá sa vyznačuje jednoduchosťou a jej základná verzia pozostáva z pätnástich elementov. Ani jeden z nich nie je povinný, ale zároveň sa môžu rozšíriť o kvalifikované elementy [25].

3.4.3 METS

Metadata Encoding and Transmission Standard (METS) je štandard metadát pre kódovanie popisných, administratívnych a štrukturálnych metadát objektov v digitálnej knižnici vyjadrených pomocou jazyka XML (Extensible Markup Language). METS je kontajnerový typ, ktorý umožňuje použitie ďalších štandardov súčasne. Vytvorené XML záznamy umožňujú popísať hierarchickú štruktúru digitálnych objektov a ukladať s nimi spojené metadáta [26].

3.4.4 EAD

Encoded Archival Description (EAD) je medzinárodný štandard metadát pre hierarchické popisy archívnych záznamov využívajúci formát XML. Schéma EAD špecifikuje prvky, ktoré sa majú použiť na popis kolekcií dát, ako aj usporiadanie týchto prvkov (napríklad ktoré prvky sú povinné, alebo ktoré sú povolené byť vo vnútri iných prvkov). Sada značiek EAD má 146 prvkov a používa sa na popis kolekcie ako celku a tiež na kódovanie podrobného viacúrovňového inventára (hierarchie). Existuje v troch verziách EAD, EAD 2002 a EAD3 [27].

3.5 Štandardizované formáty archivačných balíkov

Archivačné systémy podľa modelu OAIS vytvárajú balíky, ktoré obsahujú dáta, ktoré sa uchovávajú, a k nim potrebné metadáta rôznych štandardov a ďalšie doplňujúce informácie. Tieto všetky na seba nadväzujúce súbory musia byť ukladané spolu. A práve to, akú štruktúru budú mať tieto balíky, sa rieši pomocou špeciálnych formátov, ako je napríklad BagIT alebo E-ARK.

3.5.1 BagIt

Je hierarchický formát balenia súborov určený na ukladanie a prenos ľubovoľného digitálneho obsahu. Takýto balík (bag) pozostáva z adresára, ktorý obsahuje hlavné ukladané (payload) dáta a prináležiace sprievodné metadáta, označované tiež ako tagovacie súbory. Tie sa používajú na zdokumentovanie obsahu a prenosu balíka.

Ten musí obsahovať aspoň jeden súbor (manifest), ktorý uvádza ukladané súbory a kontrolné súčty podľa deklarovaného algoritmu, viď ukážka obsahu 3.5.1. Zabezpečuje tak validitu dát, keď v ňom uložené odtlačky sa porovnávajú pri procese archivácie s aktuálnymi hodnotami, a zisťuje sa či nebol obsah balíku zmenený, poškodený alebo nahradený. Všetky ostatné metadáta sú voliteľné, ale ich obsah musí spĺňať kritéria syntaxu popísané v špecifikácií [28].

3.5.1: Ukážka obsahu manifest súboru.

```
0d8087a24bd227db738676a4b1073242 data/diplomova-praca/schema.png
1613f3eeec0628c527a7dd80e71771e5 data/diplomova-praca/graf.jpg
18366ff278caa580bf56ddcaea8f1c4f data/diplomova-praca/dokument.pdf
0e9dd0e0683c984846ec69435800ec01 data/diplomova-praca/popis.txt
```

Pre potreby prenosu je zvyčajne síce štruktúra BagIt zabalená do formátu ZIP alebo TAR ale samostatne sa od nich líši hlavne tým, že poskytuje silné záruky integrity podporou hashovacích algoritmov ako MD5, SHA1, SHA256, SHA512 (SHA512 ako základ, ostatné podporované pre spätnú kompatibilitu). Taktiež poskytuje priamy prístup k súborom, kedy netreba žiadne špeciálne nástroje na prácu s obsahom, takže odpadá spracovanie potencionálne veľkých archívnych súborov a extrakcia potrebných častí dát. BagIT nemá stanovenú maximálnu veľkosť individuálnych súborov alebo celého balíka, preto je aj vhodný pre digitálnu archiváciu.

Aby bola zaručená interoperabilita, musia sa pri implementovaní BagIt nástrojov brať do úvahy rozdiely jednotlivých platforiem, operačných systémov, atď., kde ide hlavne o spôsob zapisovania cesty súborov, rezervované názvy, maximálna dĺžka cesty a podobné [29].

3.5.2 E-ARK

E-ARK (European Archival Records and Knowledge Preservation) bol nadnárodný výskumný projekt (rok 2014 – 2017), ktorý zdokonalil metódy a technológie digitálnej archivácie s cieľom dosiahnuť konzistenciu v celoeurópskom meradle. Riešením radu problémov spojených s nezávislými technológiami, systémami a postupmi pri vedení záznamov bol E-ARK projekt prospešný pre rozvoj medzinárodne prístupných archívov prostredníctvom poskytovania technických špecifikácií a nástrojov, rozvoja integrovanej infraštruktúry pre archiváciu, a ďalšie [30].

Všetky nástroje, aplikácie, systémy, ktoré vzišli z tohto projektu majú otvorený zdrojový kód. Ide o RODA-in, RODA, ESSArch, E-ARK Web a iné. Dôležitým bolo aj ustanovenie štandardu pre SIP, pretože síce referenčný model OAIS popisuje tieto balíky, ale nikde nešpecifikuje ako majú vyzerať. E-ARK verzie 1 bol upravený novým projektom E-ARK4ALL (rok 2018 – 2019) na verziu 2, ktorý vylepšil a doplnil

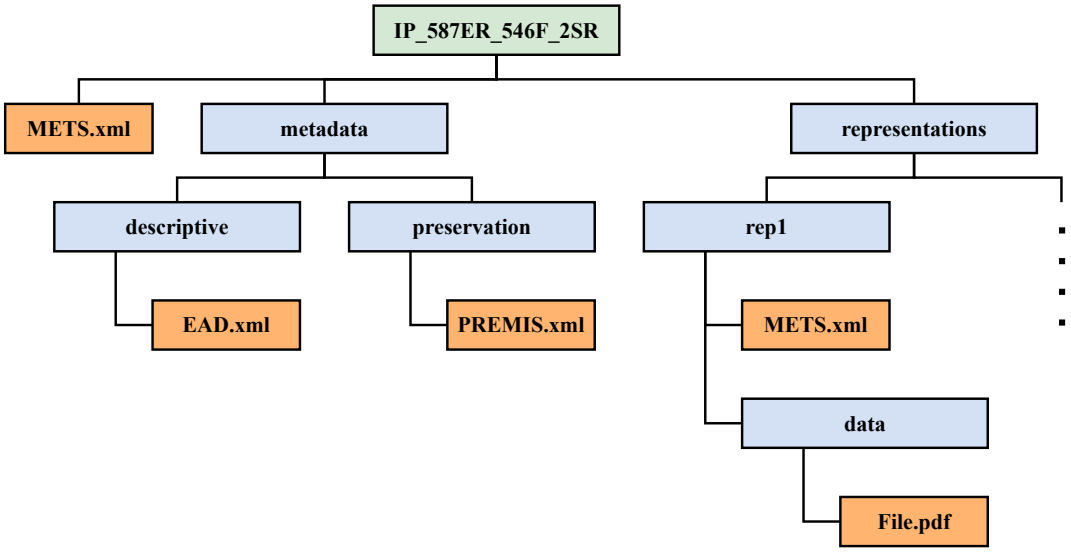
štandard na základe prevádzky a potreby viacerých európskych národných archívov. Jeho cieľom bolo znížiť náklady na dlhodobú dostupnosť dát. Aktuálne prebieha tretí európsky projekt E-ARK3, ktorého výsledky budú zverejnené koncom roka 2021. Ten sa sústredí na priblíženie systému dlhodobej archivácie malým organizáciám, podnikom v súkromnom sektore, ako aj samosprávam či jednotlivcom [31].

Komplexný popis informačných balíkov SIP, AIP a DIP a ich tvorby existuje v dokumente CSIP (Common Specification for Information Packages) [32].

Štruktúra balíka E-ARK SIP

Balík SIP môže prejsť kompresiou napríklad TAR alebo ZIP. Štruktúra je nasledovná [33]:

- Obsah každého informačného balíka musí byť uložený v jednom samostatnom priečinku, ktorý by mal byť pomenovaný podľa ID balíka, alebo ako ID plus názov, či názov a čas vytvorenia, viď obr. 3.2.
- Musí obsahovať súbor METS.xml s metadátami o identite a štruktúre balíka a jeho komponentov.
- Ďalej je potrebná zložka metadata obsahujúca všetky relevantné metadáta balíka, ktorá sa môže ďalej deliť na podzložky pre popisné (descriptive) a archivačné (preservation).
- Ďalšou povinnou zložkou je representations, ktorá obsahuje všetky archivované dáta a súbor METS.xml obsahujúci informácie o identite a štruktúre daných reprezentantov.
- Voliteľným priečinkom je schemas, ktorý obsahuje jednotlivé schémy o metadátach, a documentation, v ktorom môže byť napríklad používateľský manuál.



Obr. 3.2: Štruktúra jednoduchého E-ARK SIP balíka.

4 Zabezpečenie ochrany dát

Táto kapitola zahŕňa popis a porovnanie médií pre ukladanie údajov, súborový systém Btrfs ako aj možnosti redundancie dát a celodiskového šifrovania. Taktiež popisuje vybrané často využívané kryptografické algoritmy, štandardy a techniky v archivačných systémoch a privátnych cloudoch pri ich zabezpečení a chode. Na záver rozoberá problematiku digitálnych podpisov.

4.1 Fyzická bezpečnosť

Dlhodobá archivácia vyžaduje potrebnú analýzu vhodných hardvérových médií (nosičov) a následne zvolenie vhodnej starostlivosti po celú dobu uloženia dát. Elektronické nosiče majú obmedzenú dobu životnosti. Okrem fyzikálnych vlastností, ktoré umožňujú životnosť od niekoľko až po desiatky rokov, ide aj o vývoj technológií, kedy sú zariadenia a programy postupom času zastarané. Predpokladá sa, že média sa budú počas archivácie meniť. Medzi hlavné kritéria výberu sa radia napríklad životnosť, rýchlosť zápisu, čítania a vyhľadávania, kapacita, predpoklad zastarania technológie ako aj cena (obstarávacia cena aj prevádzka). Preto je dôležité vytvoriť vhodnú architektúru uloženia, na základe potreby a typu uchovávaných dát, kde výber vhodného média, spôsob kontroly a údržby, či myslenie na starnutie technológií hrá veľkú rolu pri bezpečnosti dlhodobej archivácie [34], [35].

4.1.1 Pevné disky

V súčasnosti najrozšírenejšie médium pre ukladanie dát. Z toho dôvodu ide aj o celkom lacné riešenie pre ukladanie dát. Pevné disky sú dostupné s rôznymi úložnými kapacitami (až niekoľko TB) a rýchlosťami. Základnú technológiu zastupujú jedna alebo viacero pevných platní s magnetickým povrchom, ktoré sa dokážu otáčať veľkou rýchlosťou, zvyčajne 5400 až 15 000 otáčok za minútu. Na tieto platne sú dáta nahrávané, mazané prostredníctvom magnetizácie feromagnetického materiálu. O tento proces sa starajú čítacie/zapisovacie hlavy.

Práve pre veľa pohyblivých častí, je priemerná životnosť takéhoto disku približne okolo 5 rokov. Záleží ale na jeho používaní a údržbe. Ak sú na disku uchovávané dáta, ku ktorým sa pristupuje len málokedy, životnosť sa kľudne môže predĺžiť na 10 a viac rokov, ale miera chybovosti sa po 5. roku rapídne zvyšuje. Naopak, ak sa s diskom stále pracuje, alebo je v nedobre vetranom zariadení (vysoké teploty), tak už po druhom roku prevádzky sa môže vyskytnúť porucha.

Miera chybovosti závisí aj od počtu pohyblivých častí v jednom disku, ako aj od rýchlosti platní. Pre archivačné účely sa využívajú priemerne kapacitné disky,

s nižšími rýchlosťami, aby vytvárali menej tepla a mali menšiu mieru chybovosti motora či samostatných hlavičiek. Taktiež je časté zrkadlenie takýchto diskov, aby sa predišlo náhlej strate dát [35].

Výhody:

- rýchly náhodný prístup k údajom,
- kompatibilita so širokou škálou platforiem a zariadení,
- možnosť prístupu k súborom od viacerých používateľov zároveň.

Nevýhody:

- vysoká spotreba energií,
- relatívna hlučnosť a vytváranie tepla,
- nízka životnosť (priemerne 5 rokov).

4.1.2 Solid State Drives

SSD alebo tiež Flash zariadenia, je technológia ukladacích médií, ktorá neobsahuje žiadne pohyblivé časti. Je založená na polovodičových čipoch. SSD dosahuje vysoké rýchlosti prístupu k uloženým údajom pri nízkej spotrebe energií. Použité čipy majú ale obmedzený počet prepisovania údajov, takže čím viac sa s dátami manipuluje tým skôr sa vyskytnú chyby. Priemerná životnosť je približne 5 až 7 rokov. Oproti pevným diskom sú ale oveľa menšie hlavne v porovnaní s tými veľkokapacitnými.

V prostredí archivácie sa moc nevyužívajú hlavne pre ich vysokú cenu a fakt, že ich najväčšia výhoda, rýchlosť, nie je pri dlhodobom uchovávaní prioritou. Ak ale cena nebude hrať rolu, ide o vynikajúce médium pre archiváciu, pretože pri dlhodobom uchovávaní sa dáta často neprepisujú, preto životnosť dokáže byť omnoho dlhšia. Tak isto miera chybovosti sa dá dobre detekovať, pretože pamäťové čipy degradujú postupne, a tak používateľ vie odhadnúť potrebu výmeny [35].

Výhody:

- nízka spotreba energií,
- kompaktná veľkosť,
- veľmi rýchly náhodný prístup k dátam,
- relatívne dobrá detekcia chybovosti.

Nevýhody:

- vysoká cena.

4.1.3 Optické disky

Najznámejšie optické disky sú CD, DVD a Blu-ray. Každý z nich má trochu odlišnú technológiu ale podstata je rovnaká. Existujú dva základné typy a to neprepisovateľné, kde sú dáta vypálené laserom do disku bez možnosti následnej zmeny,

a prepisovateľné, kde je špeciálna vrstva len nahrievaná a chladená za účelom uchovania informácií. Tento proces je obmedzený v počte opakovaní (pri Blu-ray približne 10 000 krát). Kapacita je pri CD 700 MB, pri DVD do 17 GB a trojvrstvové Blu-ray dokáže uložiť až 128 GB.

Pri dlhodobej archivácií sa optické disky využívajú ojedinele a skôr pri malom objeme dát, pretože je nepraktické z časového aj organizačného hľadiska ukladať dáta na tak malé médiá. Tieto disky sú relatívne lacné, ale pri potrebe väčšej kapacity už cena nie je priaznivá. Výhodou je ich životnosť, dokážu zachovať dáta až 50 rokov. Táto doba ale veľmi závisí od kvality média, zapisovacieho zariadenia, počtu čítaní dát a hlavne skladovacích podmienok týchto optických diskov. Z pohľadu dlhodobej archivácie je neperspektívne využitie tohto typu uloženia informácií, pretože táto technológia je v súčasnej dobe na ústupe a nie je istá jej budúcnosť [35].

Výhody:

- dlhá životnosť napálených údajov.

Nevýhody:

- malá kapacita,
- otázna podpora zariadení v budúcnosti,
- žiadna alebo obmedzená možnosť prepísania dát.

4.1.4 Dátové pásky

Technológia je založená na ukladaní dát na magnetické pásky. Od svojho vzniku sa používa množstvo rôznych typov pások, niektoré sú už zastarané (napr. DAT pásky). V dnešnej dobe sú magnetické dátové pásky jednou z najbežnejšie používaných hardvérových úložných médií pre digitálne súbory vo väčších archívoch v kombinácií s pevnými diskami. Najrozšírenejším typom sú jedno-nábojové pásky LTO (Linear Tape Open), ktoré existujú do veľkosti niekoľko desiatok TB. Novšie pásky (od verzie LTO-4) majú podporu šifrovania, a taktiež verziu WORM (Write Once Read Many), kedy sa zabraňuje prepísaniu uložených dát. Aj keď sa jedná o staršiu technológiu, stále sa aktívne vyvíja pre hlavne podnikové prostredie, kde sa neplánuje od jej používania zatiaľ upustiť. Pásky, z pohľadu ich konštrukcie sú náchylné na poškodenie, ak sa ich obsah často číta alebo zapisuje, pretože sa vždy musí navinúť na správne miesto, čo trvá aj istý čas (táto vlastnosť sa neustále vylepšuje). Preto sa odporúča, bez ohľadu na to, aká spoľahlivá a kvalitná páska je, v prostredí s častým prístupom k dátam používať pevné disky. Cena pásky je v porovnaní s klasickým diskom nižšia, ale náklady sa zvyšujú potrebou zakúpenia špecializovaného zariadenia na ich čítanie a zápis.

Dátové pásky dokážu uchovať informácie až 30 rokov, preto sú veľmi vhodné na dlhodobú archiváciu. Problémom je, že každé 3 roky vychádza nová generácia

pások a taktiež zariadení a softvéru potrebného na prácu s ich obsahom. Aj keď sa dbá na zabezpečenie spätnej kompatibility pre staršie verzie, odporúča sa po 3 generáciach prejsť na tú najnovšiu verziu média (čo zodpovedá približne každých 10 rokov). Taktiež, ak sa páska nevyužíva, je potrebné aspoň raz ročne skontrolovať jej stav, a všetky dáta na nej uložené [36].

Výhody:

- nízka cena,
- nenáročnosť na energie,
- nízka chybovosť v porovnaní s pevnými diskami,
- životnosť až do 30 rokov,
- podpora kompresie a šifrovania.

Nevýhody:

- pomalý náhodný prístup k dátam,
- veľmi drahé pri použití automatických robotických meničov a uskladňovačov,
- staré zariadenia nie sú kompatibilné s novými páskami.

4.1.5 Zhrnutie fyzickej bezpečnosti

V súčasnej dobe sú pevné disky najpoužívanejším a najdostupnejším spôsobom na uchovávanie veľkého objemu dát. V budúcnosti budú nahradené SSD, ak sa ich cena dostatočne zníži. Už teraz prebieha obmena vo väčších datacentrách, kde sa to vyplatí kvôli úspore energií a lepším vlastnostiam SSD. Doplnkom sú dátové pásky, ktoré majú dobrú vlastnosť pri dlhodobom uchovávaní dát. Problémom je ich nedostupnosť pre jednotlivcov a menšie firmy pre drahé čítacie zariadenia. Optické disky sú na dlhodobú archiváciu nevhodné pre ich malú kapacitu a hlavne neistú podporu do budúcnosti (upúšťa sa od tejto technológie). Bez ohľadu na to, aká stratégia a typ médiá sa vyberie pre archiváciu, treba zvážiť zahrnutie do plánu redundanciu na inom type úložiska. Tento prístup významne minimalizuje riziko výberu nesprávneho dlhodobého média.

4.2 Softvérová bezpečnosť

Táto podkapitola približuje súborový systém Btrfs, techniky zabezpečenie redundancie dát ako aj celodiskové šifrovanie pomocou LUKS.

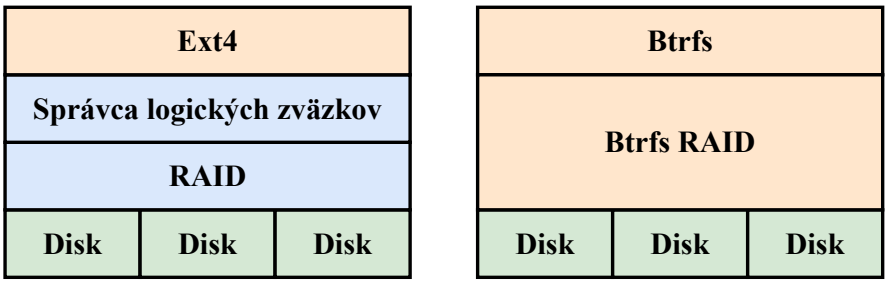
4.2.1 Btrfs

Btrfs je moderný súborový systém typu CoW (copy on write) neustále vyvíjaný pod licenciou GPL, zameraný na implementáciu pokročilých funkcií, odolnosti voči

chybám a ich ľahkú opravu. Jednoduchá správa je tiež jedným z cieľov [37].

Najdôležitejšie funkcie Btrfs:

- **Snapshots** – udržiavajú stav dát k určitému okamžiku, ktoré je možné použiť k zálohovaniu. Nevytvárajú úplnú kópiu dát, ale zaznamenávajú len zmeny a tým šetria miesto na úložisku.
- **CoW** – zaručuje konzistenciu dát, pretože pokiaľ nie je dokončená operácia, vždy existujú pôvodné dáta.
- **Detekcia chýb** – je zabezpečená vytváraním kontrolných súčtov dát aj metadát, čím je možné už na úrovni súborového systému zistiť chybu.
- **Automatická oprava** – detekcia chýb pomocou kontrolných súčtov umožňuje napríklad v prípade RAID 1 automaticky neporušené dáta načítať z druhého disku, a následne ich na prvom opraviť.
- **RAID** – podpora softvérového RAID 0, RAID 1, RAID 10.
- **Zmeny za chodu** – všetky zmeny počas behu systému je možné vykonať bez nutnosti reštartu. Napríklad zmena veľkosti oddielov, pridanie disku, zavedenie zrkadlenia či vytváranie a práca so snapshotmi.



Obr. 4.1: Porovnanie Ext4 a Btrfs RAIDu.

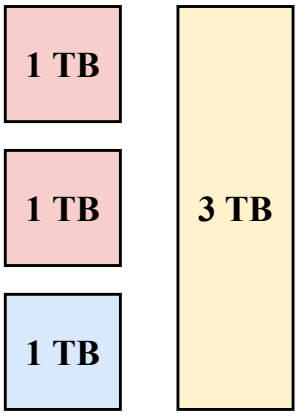
Súborový systém Btrfs nepodporuje použitie RAID 5 a 6, respektíve našli sa v implementácií chyby, ktoré ešte nie sú úplne opravené, preto sa tieto typy neodporúča používať. Taktiež šifrovanie na úrovni súborového systému je vo vývoji, preto je zatiaľ jedinou možnosťou použitie technológie LUKS nad Btrfs [38].

V porovnaní s klasickým súborovým systémom Ext4, Btrfs má veľa funkcií implementovaných, viď obr. 4.1, a tak nie je potreba používať dodatočné nástroje tretej strany. Ide napríklad vstavané funkcie:

- kompresia,
- deduplikácia,
- RAID,
- variabilná veľkosť bloku súboru.

Výhodou zrkadlenia Btrfs, je aj možnosť kombinácie diskov rôznych veľkostí, viď obr. 4.2. Je možné teda použiť ľubovoľné disky, ktoré sú k dispozícií. Príkladom je

RAID 1 zložený z hlavného 3 TB disku, ktorý sa zrkadlí na tri 1 TB disky. Taktiež je možné zrkadliť zo začiatku na jeden menší disk, a postupne podľa obsadenosti hlavného pripojiť ďalšie podľa potreby [37].



Obr. 4.2: Možnosť rozdelenia diskov pri Btrfs RAID 1.

4.2.2 Redundancia dát

RAID (Redundant Array of Independent Disks) sa označuje metóda zabezpečenia dát proti zlyhaniu disku. Je realizované hardvérovými radičmi alebo softvérovo, špecifickým ukladaním dát na viac nezávislých diskov. RAID ale nemožno považovať za zálohovanie dát, ale skôr ako redundanciu hardvéru kvôli zlyhaniu. Najčastejšie sa používa:

- **RAID 0** – nedá sa považovať za čistokrvný RAID, pretože neobsahuje žiadne redundantné informácie. Jednotlivé disky sú spojené do logického celku a neposkytujú tak dátam žiadnu ochranu. Toto spojenie môže byť lineárne alebo prekladaním, ktoré môže zrýchliť čítanie a zápis väčších blokov dát.
- **RAID 1** – klasické zrkadlenie diskov, kedy sa obsah zaznamenáva na 2 naraz. V prípade zlyhania jedného sa pracuje s kópiou. Predstavuje vysokú bezpečnosť dát, ale nižšiu rýchlosť.
- **RAID 5** – potrebuje pre svoj chod aspoň 3 pevné disky. Na dvoch sú uložené súbory a na treťom samo opravné kódy. V prípade zlyhania jedného disku sa dajú dáta obnoviť, pri 2 už nie. Takéto riešenie prináša vyššiu rýchlosť čítania dát ale pomalšie zapisovanie.
- **RAID 6** – podobný RAID 5, ale s jedným diskom naviac. Výpočet kontrolných dát je robený spôsobom aby bolo možné obnoviť dáta aj po strate 2 diskov.
- **RAID 10** – vznikol kombináciou RAID 1 a 0. Najprv sa disky zrkadlia a následne je na každý zväzok využitý princíp z RAID 0. Výhodou je tak vysoká rýchlosť zápisu a čítania, ako aj bezpečnosť dát.

Všeobecne sa neodporúča pre RAID kupovať disky z jednej výrobnej série. Vo veľa prípadoch sa ukázalo, že takéto disky zvyknú zlyhávať v podobnú dobu, a preto je väčšia pravdepodobnosť úplnej straty dát [39].

4.2.3 LUKS

Linux Unified Key Setup je štandard pre šifrovanie celých diskov, nezávislý na platforme, bežne implementovaný do rôznych operačných systémov založených na Linuxe. Existuje v dvoch verziách. Základný princíp fungovania je taký, že sa zašifruje na danom disku každý bit, takže keď sa médium dostane do nesprávnych rúk, bez znalosti hesla (bezpečnostnej frázy) sa k údajom nikto nedostane.

Najväčším problémom pri celodiskových šifrovacích riešeniach je správa hesiel. Používatelia zvyčajne vyberajú slabé heslá pretože sú ľahšie zapamätateľné. Tento problém rieši LUKS implementáciou dvojúrovňovej kľúčovej hierarchie. Hlavný kľúč je chránený derivačnou funkciou PBKDF2 (postupne nahrádzané Argon2). Základnou šifrou je AES s módom XTS a veľkosťou kľúča 256 bitov. Pre dosiahnutie najlepšieho výkonu šifry sa odporúča použitie zariadení s procesormi, ktoré podporujú inštrukcie AES-NI [40].

LUKS sa nastavuje pomocou nástroja príkazového riadku Cryptsetup, ktorý ovláda modul dm-crypt pre vytvorenie, pristupovanie a manažovanie šifrovaných zariadení. Po procese vytvorenia a šifrovania disku podľa LUKS, je označovaný ako uzamknutý. Na jeho odomknutie existujú viaceré možnosti:

- **Manuálne odomknutie** – po spustení zariadenia LUKS disky zostanú zašifrované. Až po manuálnom zadaní potrebného príkazu a následnej bezpečnostnej frázy, bude obsah sprístupnený.
- **Odomknutie pri spustení** – pri spúšťaní systému, je používateľ vyzvaný na zadanie bezpečnostnej frázy pre odomknutie daných diskov.
- **Automatické odomknutie** – vytvorí sa kľúčový súbor, ktorý je uložený na systémovom disku. Po spustení systému sa kľúč z tohto súboru automaticky načíta a odomkne LUKS disk. Existuje viacero typov kľúčových súborov:
  - Kľúčový súbor s uloženou jednoduchou bezpečnostnou frázou.
  - Kľúč v súbore je náhodne vygenerovaný reťazec znakov.
  - Za kľúčový súbor je označený ľubovoľný súbor so statickým binárnym obsahom (napríklad obrázok, video, atď.), ktorý znemožňuje priamu detekciu útočníkom, že ide o kľúčový súbor.

LUKS podporuje šifry rodiny AES (AES, Twofish, Serpent, atď.), v módoch CBC, XTS s veľkosťami kľúča 256 a 512 bitov. Podporované hashovacie algoritmy sú SHA256 a SHA512. Na odvodzovanie kľúča zo zadanej frázy sa vo verzií LUKS 1 používa PBKDF2, a v LUKS 2 je nahradený Argon2 (preferovaný Argon2i).

LUKS verzie 2 vylepšuje a doplňuje základnú verziu. Je ale, dostupná len na novších verziách operačných systémov (nutná podpora Cryptsetup 2.0) . Taktiež je náročnejší na výpočtový výkon, hlavne kvôli použitiu Argon2, a taktiež zvýšeniu základného iteračného času pre proces odvodzovania kľúča na 2 sekundy. Celkovo je druhá verzia považovaná za veľmi bezpečnú, nie sú zatiaľ známe žiadne závažné bezpečnostné hrozby [41].

4.3 Kryptografická bezpečnosť

Táto kapitola popisuje kryptografické šifrovacie algoritmy ako AES s vybranými módmi. Ďalej rozoberá problematiku funkcie hash ako aj digitálnych podpisov podľa rôznych štandardov.

4.3.1 AES

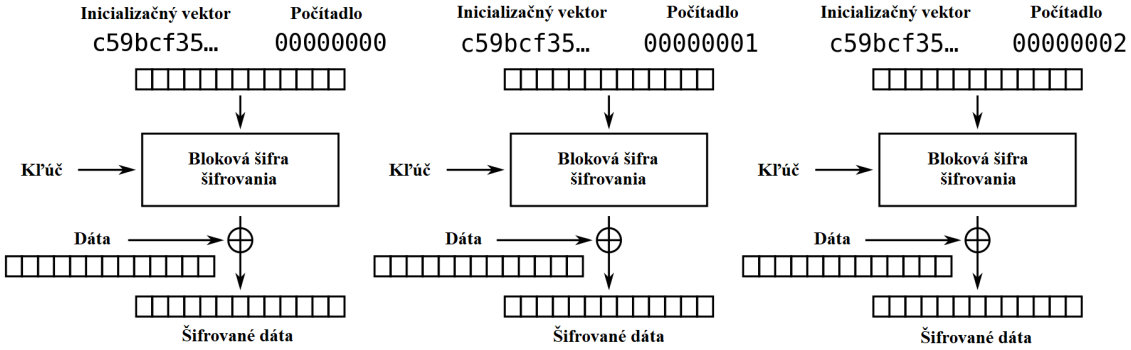
Advanced Encryption Standard (AES), je symetrická bloková šifra tiež známa ako Rijndael, ktorá je štandardom od roku 2001. Spracovávané dáta rozdeľuje do blokov fixnej dĺžky 128 bitov. Kľúč môže mať veľkosť 128, 192 a 256 bitov, ktorý sa používa na šifrovanie aj dešifrovanie. Algoritmus preženie dáta niekoľkými okruhmi s rôznymi základnými operáciami. Počet kôl sa odvíja od použitej dĺžky kľúča, AES-128 má 10 kôl, AES-192 12 kôl a AES-256 14 kôl.Každé jedno z nich je zložené zo 4 operácií: AddRoundKey (pripočítanie podkľúča), SubBytes (substitúcia bajtov), ShiftRows (posun riadkov), MixColumns (miešanie stĺpcov) [42], [43].

Symetrická bloková šifra AES môže pracovať v rôznych módoch, ktoré sa delia do dvoch základných skupín [44]:

- poskytujúce dôvernosť,
  - ECB (Electronic Code Book),
  - CBC (Cipher Block Chaining),
  - OFB (Output FeedBack),
  - CTR (Counter Mode),
  - XTS (XEX Tweakable Block Ciphertext Stealing),
- poskytujúce dôvernosť a autenticitu,
  - CCM (Counter Mode Cipher Block Chaining Message Authentication Code Protocol),
  - GCM (Galois/Counter Mode),
  - CWC (Carter-Wegman CTR mode),
  - OCB (Offset Codebook Mode).

CTR

CTR mód šifruje dáta postupne po jednom za sebou (v jednotlivých blokoch). Každý blok obsahuje nový nepredikovateľný reťazec dát, ktorý je generovaný z inicializačného vektora s počítadlom spolu s šifrovacím kľúčom, viď obr. 4.3. Následne je pomocou operácie XOR zlúčený so vstupnými dátami. Výsledkom sú šifrované dáta. Vstupné údaje (pred šifrovaním) a výstupné údaje (po šifrovaní) majú rovnakú dĺžku. Mód CTR je vo väčšine prípadov považovaný za dobrú voľbu kvôli silnému zabezpečeniu, ľubovoľnej dĺžke vstupných údajov (bez potreby výplne) a možnostiam paralelného spracovania (dobrá rýchlosť) [42].



Obr. 4.3: CTR mód [42].

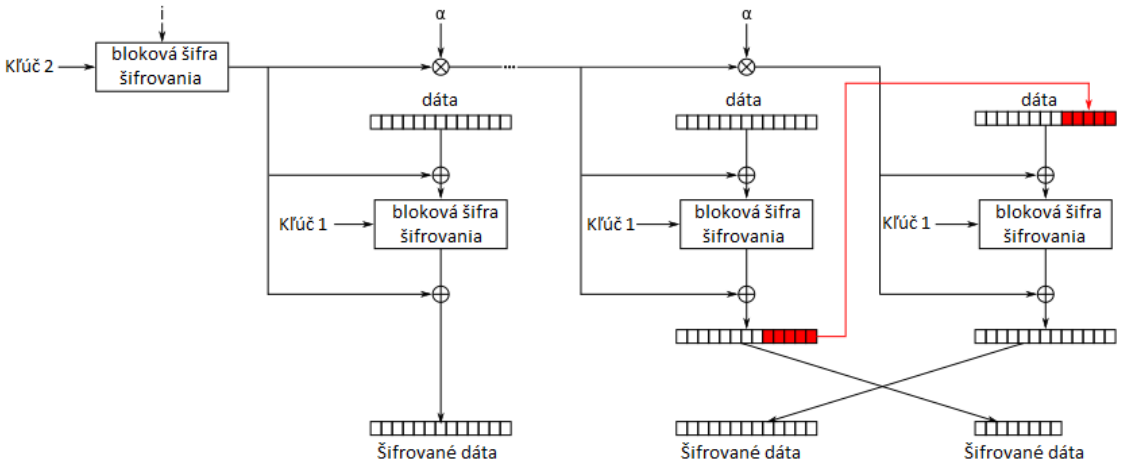
GCM

Mód GCM využíva všetky výhody CTR a pridáva autentifikáciu správ (vytvára kryptografický overovací tag správy – MAC message authentication tag). Mód GCM je rýchly a efektívny spôsob implementácie autentizačného šifrovania v symetrických šifrách a všeobecne sa veľmi odporúča jeho použitie v tomto smere [44].

XTS

Tento mód je založený na XEX (XOR Encrypt XOR). XTS využíva dva kľúče, kde prvý je určený na klasické šifrovanie bloku AES, a druhý na šifrovanie takzvanej vylepšenej hodnoty (Tweak Value), viď obr. 4.4. Využíva metódy krádeže šifrovaného textu (ciphertext stealing). Týmto sa dosiahne to, že každý blok produkuje jedinečný šifrový text, ktorý má identický základ, bez použitia inicializačných vektorov a reťazenia. Text je tak v skutočnosti takmer (nie celkom) šifrovaný dvakrát pomocou dvoch nezávislých kľúčov. Tento mód je najvhodnejší režim pre šifrovanie

celého disku, ale keďže je náchylný na manipuláciu a falšovanie dát a klasické súborové systémy ako ext4 a NTFS nemajú proti tomuto typu útoku ochranu, odporúča sa použitie v kombinácií so ZFS alebo Btrfs [45].



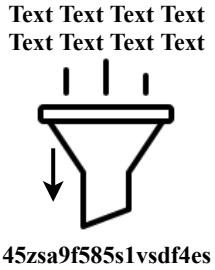
Obr. 4.4: XTS mód.

AES-NI

Na vybraných zariadeniach (procesoroch) sú dostupné takzvané inštrukcie novej generácie šifry AES. Ich hlavnou úlohou je šifru zefektívniť, zrýchliť a naplno využiť hardvér pre spracovanie potrebných požiadaviek. Druhotným efektom je zlepšenie bezpečnosti, kvôli eliminácií veľkej časti časových útokov, alebo útokov na vyrovnávaciu pamäť. Tým je možné zavedenie vyššieho štandardu bezpečnosti pre AES (väčšia veľkosť kľúčov), bez nutnosti obmedzovania výkonu pre ostatné úlohy, pretože nové inštrukcie dokážu zvýšiť výkon šifry 3 až 10 násobne oproti klasickým softvérovým riešeniam [46].

4.3.2 Funkcia Hash

Pod pojmom funkcia Hash sa rozumie mapovanie textu (dát) do jedného reťazca, viď obr. 4.5. V kryptografickom ponímaní sa z dát vytvorí odtlačok fixnej dĺžky, kde je tento proces nenávratný a odolný proti kolíziám. Nenávratnosť znamená, že výsledný hash sa nedá dostať naspäť do podoby vstupných dát. Odolnosť voči kolíziám zase, že rozdielny vstup nevyústi v rovnaký výstup. Už malá zmena na vstupe sa odrazí vo veľkej na výstupe. V podobe kryptografickej funkcie sa využíva pre zaistenie integrity dát, ochranu uložených hesiel, vytváranie a overovanie elektronického podpisu a podobne. Najpoužívanejšia funkcia v tejto dobe je SHA-256 z rodiny SHA-2, ktorá sa považuje za dostatočne bezpečnú [47].



Obr. 4.5: Hash funkcia.

Fixity

V prostredí uchovávania dát tento výraz predstavuje zabezpečenie, že digitálny súbor zostane nezmenený (opravený). Táto oprava sa nevzťahuje len na súbory, ale na akýkoľvek digitálny objekt, ktorého štruktúra (séria bitov) musí zostať zachovaná, neporušená. Funkcia Fixity môže byť aplikovaná na obrázky alebo video v audiovizuálnom objekte, taktiež na súbory vo vnútri zip balíka, metadáta v XML štruktúre, jednotlivé záznamy v databáze, alebo na objekty v archívnom úložisku. Monitoring neporušenosti prebieha za pomoci kontrolných súčtov (checksums), kde sa takzvaným digitálnym odtlačkom súboru zabezpečí, že aj tá najmenšia zmena bude detekovateľná [48]. Kontrolné súčty sa používajú keď treba zabezpečiť:

- Detekciu správneho procesu prijatia súboru od vlastníka obsahu alebo zdroja a úspešné prenesenie do archivačného úložiska.
- Vedomosť, že funkcia Fixity bola vykonaná pri ukladaní súboru.
- Správnu extrakciu z archívneho úložiska a doručenie dát v neporušenom stave používateľom v budúcnosti.

Pretože ide o pomerne jednoduché funkcie, kontrolné súčty sú integrované do mnohých nástrojov na digitálne uchovávanie. Existuje niekoľko rôznych algoritmov kontrolného súčtu, napríklad MD5, SHA-256. Čím je algoritmus „silnejší“, tým ťažšie je zámerne zmeniť súbor spôsobom, ktorý zostane nezistený. Táto vlastnosť môže byť dôležitá, keď je potrebné preukázať čo najlepšiu odolnosť systému, napríklad v prípade, že by mohli byť pozmenené citlivé alebo tajné materiály. Ak sa však kontrolným súčtom zisťujú len náhodné straty či poškodenia súborov v dôsledku zlyhania úložného média, potom majú veľkú výhodu práve ľahšie vypočítateľné algoritmy ako MD5 [49].

4.3.3 Doporučené dĺžky kľúčov

Vo väčšine kryptografických funkcií je dĺžka kľúča dôležitým bezpečnostným parametrom. Existujú viaceré známe organizácie, ktoré vyhodnocujú minimálne bezpeč-

nostné požiadavky pre dané algoritmy. Uvedené dĺžky kľúčov sú navrhnuté tak, aby odolávali matematickým útokom (neberú do úvahy chyby algoritmov alebo hardvéru) 4.1.

Tab. 4.1: Bitová dĺžka podľa NIST a ECRYPT [50], [44].

		NIST		**ECRYPT**	
Obdobie (roky)		**2019 – 2030**	**2030 +**	**2018 – 2028**	**2029 – 2068**
Symetrické šifry		112	128	128	256
Šifry využívajúce faktorizáciu		2048	3072	3072	15 360
Diskrétny logaritmus	**Kľúč**	224	256	256	512
	Skupina	2048	3072	3072	15 360
Šifry využívajúce eliptické krivky		224	256	256	512
Hash		224	256	256	512

4.3.4 Digitálny podpis

V kryptografii poskytujú digitálne podpisy autentifikáciu správ, integritu a nemožnosť vyvrátenia podpisu digitálnych dokumentov. Pracujú v kryptosystémoch využívajúcich verejný kľúč a používajú páry verejných, súkromných kľúčov, kde podpisovanie správ sa vykonáva súkromným a overovanie zodpovedajúcim verejným kľúčom.

Podpis matematicky zaručuje, že správa bola podpísaná určitým (tajným) súkromným kľúčom, ktorý zodpovedá konkrétnemu verejnému kľúču. Po podpísaní správy ju nie je možné upraviť, tak isto ani podpis, a tým sa zabezpečuje autentifikácia a integrita správy. Ktokoľvek, kto pozná verejný kľúč podpisovateľa správy, môže podpis overiť. Po podpísaní nemôže autor podpisu odmietnuť tento akt (nemožnosť vyvrátenia podpisu) [51].

DSA

DSA (Digital Signature Algorithm) je kryptograficky bezpečný štandard pre digitálne podpisy (podpisovanie správ a overenie podpisu), založený na modulárnom umocňovaní a obtiažnosti problému s diskrétnym logaritmom (discrete logarithm problem). DSA sa počíta pomocou sady parametrov, a to súkromného kľúča **x**, tajného čísla **k** (je jedinečné pre každú správu), údajov ktoré sa majú podpísať, a hashovacej funkcie. Digitálny podpis sa overuje pomocou verejného kľúča **y**, ktorý je matematicky spojený so súkromným kľúčom použitým pri podpisovaní údajov, ktoré

sa majú overiť, a rovnakej hashovacej funkcie, aká sa použila počas generovania podpisu. Základné parametre ako typ hashovacej funkcie, bitová dĺžka prvočísel použitých pri výpočte, určujú základnú bezpečnosť algoritmu [52].

AdES

Advanced Electronic Signatures (AdES) je zdokonalený elektronický podpis, ktorý spĺňa požiadavky stanovené v nariadení EÚ č. 910/2014 (eIDAS) o elektronickej identifikácii a dôveryhodných službách pre elektronické transakcie. Oproti obyčajnému elektronickému podpisu, musí zaručiť navyše že [53]:

- je jedinečne prepojený a schopný identifikovať signátora,
- je vytvorený spôsobom, ktorý umožňuje signatárovi zachovať možnosť kontroly,
- sa spája s dokumentom takým spôsobom, že je možné zistiť každú následnú zmenu údajov.

XAdES

XAdES (XML Advanced Electronic Signature) je štandard, ktorý definuje formát XML pre pokročilé elektronické podpisy, ktoré môžu zostať v platnosti po dlhú dobu [54]. Existuje niekoľko verzií daného štandardu:

- XAdES-BES (basic electronic signature) – základná forma definujúca prvky autentifikácie a ochrany celistvosti záznamov, ktorá však nezabraňuje jej existencii.
- XAdES-T (TimeStamp) – pridanie časovej pečiatky zaisťuje neodmietnutie.
- XAdES-C (Complete validation data) – nadväzuje na XAdES-T pridávaním odkazov na množinu údajov podporujúcu overenie elektronického podpisu.
- XAdES-X (eXtended validation data) – vytvára sa na základe XAdES-C pridaním časových pečiatok, aby sa znížili riziko, že by mohli byť ohrozené akékoľvek kľúče použité v reťazci certifikátov alebo v informáciách o stave odvolania.
- XAdES-X-L (eXtended validation data incorporated for the Long term) – vychádza zo štandardu XAdES-X pridaním overovacích údajov (napríklad certifikáty) pre situácie, kedy je predpoklad, že sa overovacie údaje dlhodobo nebudú uchovávať.
- XAdES-A (Archiving validation data) – rozširuje XAdES-X-L pridaním časových pečiatok pre archiváciu podpisov.

CAdES

CAdES (CMS Advanced Electronic Signatures) je sada rozšírení k podpísaným údajom CMS (Cryptographic Message Syntax). Definuje množstvo formátov elektronického podpisu vrátane tých, ktoré môžu zostať v platnosti po dlhú dobu. Patria sem dôkazy o ich platnosti, aj keď sa podpisujúca alebo overovacia strana neskôr pokúsi odoprieť (odmietne) platnosť elektronického podpisu [55].

PAdES

PAdES (PDF Advanced Electronic Signature) formuluje rovnaké funkcie ako CAdES a XAdES, ale vzťahuje sa iba na dokumenty PDF kde definuje požiadavky, ktoré musí softvér na prezeranie a úpravy dodržiavať pri použití digitálnych podpisov. Tiež definuje, ako je možné zobraziť daný podpis (na akom konkrétnom mieste na konkrétnej stránke) [56].

Časová pečiatka

Elektronická časová pečiatka je údaj v elektronickej podobe, ktorý viaže ďalšie údaje v elektronickej podobe k určitému času, čím sa preukáže, že v danom čase existovali. Časové pečiatky sa zvyčajne používajú pri zaznamenávaní udalostí (logy), v takom prípade je každá udalosť v denníku označená časovou pečiatkou. V súborových systémoch môže časová pečiatka odkazovať na uložený dátum a čas vytvorenia alebo úpravy súboru. Dôveryhodná časová pečiatka je proces bezpečného sledovania času vytvorenia a úpravy dokumentu. Dôveryhodnú autoritu TSA (Time Stamping Authority) možno použiť na preukázanie konzistencie a integrity digitálnych dôkazov v každej etape ich existencie [57].

Digitálny certifikát

Digitálne certifikáty sú digitálne záznamy používané na jednoznačnú identifikáciu subjektu (totožnosti osoby, organizácie alebo stroja). Umožňujú používateľovi overiť, komu sa certifikát vydáva, a tiež vydavateľa certifikátu. Digitálny certifikát je platný počas určitého obdobia [57].

SignServer

Je aplikácia určená na vykonávanie rôznych druhov digitálnych podpisov. Má otvorený zdrojový kód, a je dostupná aj v komunitnej edícií zadarmo. SignServer je vlastne podpisový server s modulmi pre vykonávanie podpisov podľa určitých

štandardov. Podpisové kľúče sú bezpečne uložené na serveri, ktorý podporuje implementáciu hardvérového bezpečnostného modulu (HSM). Aplikácia je pripravená na použitie v prostredí kontajnerov (predpripravený obraz vývojármi) [58].

SignServer dokáže zastávať funkcie:

- Podpisovanie kódu – MS Authenticode, Java vrátane Android APK.
- Podpisovanie dokumentov – PDF, XML, XAdES (BES a T).
- Časová pečiatka – MS Authenticode, štandard RFC 3161 kompatibilný s ETSI.
- ePassport – MRTD kompatibilný s ICAO.

5 Existujúce riešenia pre privátne cloudy

V dnešnej dobe existuje mnoho spoločností, ktoré ponúkajú softvér typu klient-server na vytvorenie privátneho cloudu na vlastnom serveri prevažne pre využívanie služieb na hosting súborov. Táto kapitola približuje a porovnáva vybrané z nich. Všetky majú otvorený zdrojový kód a sú dostupné aspoň v základnej verzii zadarmo. Taktiež ponúkajú podnikové (enterprise) riešenia, za úplatu, zvyčajne s rozšírenou funkcionalitou alebo podporou. V porovnaní je braný dôraz na funkcionalitu zabezpečenia a správy.

5.1 Nextcloud

Nextcloud je sada klient-server aplikácií s otvoreným zdrojovým kódom, pre vytváranie a používanie služieb na hosting súborov. Tento softvér je dostupný od jednotlivcov až po veľké podniky (enterprise verzie), a dá sa nainštalovať a prevádzkovať na vlastnom privátnom serveri. Umožňuje zdieľanie a synchronizáciu súborov a priečinkov so serverom, podobne ako známe komerčné riešenia (napríklad Dropbox), ale naviac ponúka miestne úložisko so silným zabezpečením a možnosťou plnej kontroly nad správou. Umožňuje prehľadnú a jednoduchú manipuláciu zo strany používateľa aj správcu pomocou priehľadného webového rozhrania, ako aj mobilných či desktopových klientov pre Windows, Mac, Linux, Android a iOS. Podľa zvoleného plánu je dostupná podpora migrácie súborov ako aj profesionálny zákaznícky servis. Rozšíriteľnosť základnej funkcionality je zabezpečená pomocou doplňujúcich aplikácií priamo od vývojárov alebo komunity. Taktiež je možné vyvinúť vlastnú aplikáciu podľa svojich potrieb.

Nextcloud zabezpečuje prenos súborov pomocou protokolu SSL/TLS a priamo na úložisku pomocou 256-bitového šifrovania AES. Ak by boli obavy, že by mohol získať útočník prístup k používateľskému heslu, existuje možnosť 2-faktorového overenia. Pri prevádzke na vlastnom serveri Nextcloud poskytuje kontrolu správcovských nastavení pomocou bezpečnostného skenu priamo od vývojárov.

Nextcloud Hub ponúka celú radu ďalších vstavaných funkcionalít okrem hostingu súborov, ktoré sa dajú využiť firmami alebo jednotlivcami pri každodennej práci. Sú to napríklad platforma na kolaboráciu pri projektoch, zdieľaný kalendár, aplikácia pre videokonferencie a iné. Preto sa považuje Nextcloud za komplexné a bezpečné riešenie pre jednotlivcov alebo firmy pri zdieľaní akýchkoľvek informácií, hlavne pre jeho jednoduchosť a dobre nastavenú platobnú politiku [59], [60].

5.2 Owncloud

Owncloud je sada klient-server aplikácií s otvoreným zdrojovým kódom veľmi podobná Nextcloudu. Taktiež ponúka inštaláciu a prevádzku na vlastnom serveri. Poskytuje prístup k údajom prostredníctvom webového rozhrania alebo mobilného, desktopového klienta a zároveň poskytuje platformu na jednoduché prezeranie, synchronizáciu a zdieľanie naprieč zariadeniami. Všetko pod možnou vlastnou správou. Prípadné rozšírenie funkcionalít je umožnené pomocou API rozhrania pre aplikácie, doplnky a akékoľvek iné úložisko. Softvérový základ je poskytovaný zadarmo, ale je obmedzený iba na jedného používateľa. Väčšina pokročilejších funkcii je obsiahnutá v platenej enterprise verzii. Podporuje inštaláciu na všetkých známych distribúciách Linuxu ako aj virtualizáciu.

Používatelia môžu taktiež využiť multi-faktorové overenie, šifrovanie obsahu pomocou 256-bitového AES alebo ochranu proti ransomware a podobne. Vývojári zabezpečujú pravidelné bezpečnostné audity svojho kódu [61], [62].

5.3 Seafile

Seafile je cloudové riešenie pre ukladanie súborov vyvíjané v jazyku C a Python. Jeho základná verzia (community edition) má otvorený zdrojový kód, ale pri enterprise verzii tomu tak už nie je. Inštalácia je možná na privátny server, ak je požadovaná plná kontrola nad správou systému. Seafile bohužiaľ obmedzuje väčšinu svojich funkcií či bezpečnostných prvkov v základnej verzii. Tak isto aj rozšíriteľnosť pomocou dostupných aplikácií nie je moc veľká. Funkcionalita je dostupná z webového rozhrania tak isto ako z ostatných platforiem či zariadení pomocou klientskych aplikácií. Dokáže sa tváriť ako virtuálny disk na serveri, ktorý rozširuje úložisko v zariadeniach a umožňuje selektívne zdieľanie súborov zabezpečené heslom a rôznymi úrovňami povolení (čítanie, zápis, atď.). Seafile má jedinečný prístup k synchronizácii malých súborov. Namiesto presunutia celého súboru ho rozdelí na kúsky, aby sa mohol presunúť rýchlejšie. Teoreticky to umožňuje presúvanie mnoho malých súborov podstatne rýchlejšie v porovnaní s inými službami.

Z bezpečnostného hľadiska je dostupné 2-faktorové overenie, šifrovanie a uzamknutie súborov, vírusový sken, AD/LDAP integrácia, a ďalšie. Všetky prenosy dát sú chránené protokolom HTTPS/TLS [63], [64].

5.4 Pydio

Pydio je platforma na správu súborov pre jednotlivcov alebo podniky s jednoduchým webovým rozhraním a podporou mobilných ako aj desktopových klientov. Inštalá-

cia na vlastný hardvér je jednoduchá a ihneď dokáže prepojiť doterajšie existujúce úložiská bez potreby migrácie. Pydio Cells, je aktuálna verzia softvéru, ktorá používa „bunky“ ako zdieľané priečinky, ktoré majú svoj vyhradený priestor s extra funkciami a vlastnými povoleniami. Pydio je samo o sebe iba jadro, ktoré beží na webovom serveri a je k nemu prístup prostredníctvom ľubovoľného prehliadača či klienta. Je ideálny pre online správu súborov a šifrovanie SSL/TLS ako aj 256-bitové AES umožňuje bezpečný prenos a uloženie dát. Enterprise verzia ponúka rozšírené možnosti správy a ďalšie vylepšenia. Oproti ostatným popisovaným riešeniam má ale obmedzenú funkcionalitu a rozšíriteľnosť [65], [66].

5.5 Porovnanie vybraných riešení pre privátne cloudy

V nasledujúcej tabuľke 5.1 sú porovnané vybrané riešenia pre privátne cloudy. Zameraná je hlavne na bezpečnostné prvky pre ukladanie dát, ako aj celý systém. Taktiež sa sa bral ohľad aj na možnosť dôkladnej, ale za to jednoduchej správy. Jednotlivé kritériá sú priblížené v priloženej legende 5.5.

LEGENDA:

A) Podrobná dokumentácia – je dôležitá pri konfigurácií a riešení problémov.
B) Podpora virtualizácie – možnosť inštalácie ako virtuálny stroj s podporou známych hypervisorov.
C) Podpora kontajnerizácie – možnosť nasadenia do prostredia ľahkej virtualizácie (kontajnerov).
D) Webové rozhranie – umožňuje používateľom pracovať so svojimi súbormi bez nutnosti inštalovať akýkoľvek dedikovaný klientský softvér.
E) Príkazový riadok – možnosť spravovať systém pomocou príkazového riadku, spúšťanie vlastných skriptov.
F) Rozšíriteľnosť – základné riešenie je rozšíriteľné o ďalšiu funkcionalitu tretích strán, alebo vlastnú.
G) Spravovanie verzií súborov – táto funkcia umožňuje návrat k starším verziám súborov.
H) Uzamknutie súboru – zabraňuje používateľom aktuálne úpravy, ktoré by mohli spôsobiť konflikty súborov. Používatelia sú na túto skutočnosť upozornení.
I) Kontrola vírusov – vírusový skener pre súbory.
J) Ochrana pred Ransomware – zabraňuje strate údajov blokovaním nahrávaných súborov od klientov, ktoré sú rozpoznané ako infikované.
K) Súborový Firewall – Správca dokáže pomocou prístupových pravidiel k súborom nastaviť prístup zo špecifických IP adries, typ povolených zariadení, veľkosť nahrávaných súborov, a iné.

Tab. 5.1: Porovnanie vybraných funkcií priblížených riešení pre privátne cloudy.

Kritérium	Owncloud	Nextcloud	Seafile	Pydio
A	áno	áno	áno	áno
B	áno	áno	áno	áno
C	áno	áno	áno	áno
D	áno	áno	áno	áno
E	áno	áno	áno	áno
F	áno	áno	áno	áno
G	áno	áno	áno	áno
H	áno	áno	iba Enterprise	áno
I	áno (ClamAV)	áno (ClamAV)	iba Enterprise (ClamAV)	iba Enterprise (ClamAV)
J	iba Enterprise	áno	nie	nie
K	iba Enterprise	áno	iba Enterprise	áno
L	áno	áno	nie	iba Enterprise
M	áno	áno	áno	áno
N	áno	áno	áno	iba Enterprise
O	áno	áno	áno	iba Enterprise
P	áno	áno	áno	iba Enterprise
Q	iba Enterprise	iba Enterprise	iba Enterprise	iba Enterprise
R	AES 256/CTR	AES 256/CTR	AES 256/CBC	AES 256/GCM
S	iba Enterprise	iba Enterprise	nie	nie
T	áno	áno	áno	áno

L) Politika hesiel – administrátori môžu definovať požiadavky na heslo používateľov (počet znakov, typ, atď.) a taktiež definovať dátum vypršania, alebo zabráneniu nastavenia rovnakého hesla.

M) Ochrana pred hrubou silou – zabraňuje útočníkom skúšanie rôznych hesiel rýchlo za sebou. Časové oneskorenie pre opakované neúspešné pokusy z rovnakej IP adresy minimalizujú šancu na úspešný útok.

N) OAuth2 – je protokol pre bezpečnú autorizáciu klientov. Jeho zavedením sa výrazne zvyšuje bezpečnosť a uľahčuje integráciu aplikácií alebo webových služieb tretích strán.

O) LDAP/Active Directory integrácia – možnosť prepojenia s protokolom ľahkého prístupu k adresáru (LDAP), pre autentifikáciu a zabezpečenie používateľov (skupín) a zodpovedajúcich atribútov.

P) Multifaktorové overenie – podpora multifaktorového alebo 2-faktorového ove-

renia zvyšuje bezpečnosť tým, že útočníkom výrazne sťažuje prístup pomocou identity jedného z legitímnych používateľov.

Q) Funkcia audit – umožňuje zaznamenávať, čo používatelia a ich správcovia robia s akými údajmi, čím odrádza od ich zneužitia.

R) Šifrovanie primárneho úložiska – šifrovanie obsahu úložiska na strane servera, správca nemôže čítať dáta používateľov.

S) Šifrovanie úložiska s podporou HSM – umožňuje generovanie a uchovávanie hlavných šifrovacích kľúčov hardvérovým bezpečnostným modulom (HSM). Je tak možné úplne zabrániť v prístupe k údajom tým, ktorí majú prístup do úložiska, pretože potrebné kľúče nie sú na ňom uložené.

T) Kontrola integrity súborov – možnosť overenia integrity zálohovaných a stiahnutých súborov vytvorením a porovnaním ich jedinečných kontrolných súčtov, aby sa zistilo prípadné poškodenie súborov.

5.5.1 Zhodnotenie analýzy vybraných riešení pre privátne cloudy

Z tabuľky kritérií existujúcich vybraných riešení sa javí ako vhodné riešenie Nextcloud. Spĺňa viaceré predpoklady na vytvorenie zabezpečeného privátneho cloudu na vlastnom zariadení. Existencia podrobnej dokumentácie a široká podpora komunity dopomáha k jednoduchšej správe a promptnejšiemu riešeniu problémov. Jednoduchá rozšíriteľnosť z radu mnohých doplňujúcich funkcií (aplikácií) radí tento systém medzi tie, ktoré sú vhodné na prispôsobenie a úpravu špecifickým požiadavkám.

6 Existujúce riešenia pre archivačné systémy

Stále narastajúce množstvo digitálnych dát dopomohlo k snahe rôznych inštitúcií a spoločností o vytvorenie archivačných systémov. Pre zabezpečenie dlhodobého chodu systému a čitateľnosti uchovávaných dát, je potrebné implementovať uznávané štandardy. Táto kapitola približuje a porovnáva vybrané riešenia archivačných systémov s otvoreným zdrojovým kódom.

6.1 Archivematica

Archivematica je integrovaná sada softvérových nástrojov s otvoreným zdrojovým kódom, ktorá umožňuje používateľom spracovávať digitálne objekty od príjmu po sprístupnenie v súlade s funkčným modelom OAIS. Implementácia je zabezpečená pomocou takzvaných mikroslužieb, kde každá z nich predstavuje jeden čiastkový krok pracovného postupu, čo vo výsledku demonštruje model funkčnosti OAIS. Používatelia monitorujú a kontrolujú tieto mikroslužby prostredníctvom ovládacích prvkov vo webovom rozhraní.

Archivematica používa na generovanie dôveryhodných, autentických, spoľahlivých a na systéme nezávislých archívnych informačných balíkov (AIP) špecifikácie metadát METS, PREMIS, Dublin Core, ako aj ďalšie uznávané štandardy. Je kompatibilná so stovkami rôznych formátov súborov. Tento systém má pokročilú správu vyhľadávania a ukladacieho priestoru, čo znamená, že je umožnené vyhľadávanie potrebných uložených archivačných balíčkov priamo cez webové rozhranie, taktiež ich následne stiahnutie vo forme kompletného AIP, alebo jednotlivých objektov. Spravovanie úložiska archivačných balíkov je zabezpečené pomocou služby Archivematica Storage Service, vrátane dvojkrokového procesu vymazania, ktorý vyžaduje odôvodnenie a schválenie na vylúčenie uloženého AIP. Medzi typy používaných úložných systémov patria lokálne systémy, cloudové úložiská ako aj špecializované úložné nástroje a služby. Používatelia môžu prijímať obsah (sprístupnenie – DIP) manuálne, alebo pomocou automatizačných nástrojov automaticky z určených zdrojových umiestnení [67], [68].

Archivematica je aktívny, dynamický projekt so širokou základňou používateľov. Vývojári neustále spolupracujú s komunitou na vylepšovaní aplikácie, kde všetky nové verzie sú prístupné každému. Tým pádom sa funkcionalita neustále rozširuje a zlepšuje. Súčasťou tohto produktu je aj podrobná dokumentácia, ktorú dopĺňa užívateľské fórum s vecnými radami ku konfigurácií či riešeniu problémov.

6.2 DAITSS

DAITSS (Dark Archive in the Sunshine State) je aplikácia na archiváciu digitálnych záznamov vyvinutá Floridským centrom pre automatizáciu knižníc (FCLA). Prvotne bola vyvíjaná a využívaná pre dlhodobé uchovávanie dát univerzitami na Floride (USA), no neskôr bola sprístupnená pomocou otvoreného zdrojového kódu aj iným subjektom [69].

DAITSS poskytuje automatizovanú podporu pre jednotlivé kroky v súlade s funkčným modelom OAIS. Je navrhnutý ako sada webových služieb a mikroslužieb, ktoré vyžadujú prísnu kontrolu a zabezpečenie integrity aj autenticity archivovaného obsahu. Poskytuje stratégie aktívneho uchovávania založeného na špecifickom spracovaní pre daný formát dát. Podporuje jednotlivé formáty metadát podľa špecifikácií ako PREMIS, METS a iné [70], [71].

Keďže má DAITSS otvorený zdrojový kód a splňuje požiadavky medzinárodne uznávaných štandardov, je teoreticky možná jeho široká implementácia a úprava pre aktuálne potreby používateľov. Bohužiaľ pre chýbajúcu podrobnú dokumentáciu a podporu zo strany vývojárov je reálna implementácia dosť sťažená, v niektorých prípadoch aj nemožná. Túto skutočnosť nezlepšuje ani fakt, že pre úzku implementáciu iba medzi pár univerzitami neexistuje ani aktívna komunita používateľov.

6.3 RODA

RODA (Repository of Authentic Digital Records) je riešenie dlhodobého digitálneho úložiska, ktoré zabezpečuje chod pre všetky hlavné časti funkčného modelu OAIS. RODA je schopná prijímať, spravovať a poskytovať prístup k rôznym typom digitálneho obsahu. Je vyvíjaná ako aplikácia s otvoreným zdrojovým kódom, ktorej technológie sú podporované štandardmi OAIS, METS, Dublin Core, PREMIS a iné. Pri príjme dát nielenže overuje štandardizované SIP, ale kontroluje aj jeho obsah, či neobsahuje vírusy, spracováva identifikáciu formátov súborov, extrahuje technické metadáta a migruje formáty súborov do lepšie uchovávateľných alternatív.

RODA tiež poskytuje prístup k digitálnym informáciám v niekoľkých formách, ako je napríklad vyhľadávanie a prehliadanie prostredníctvom grafických používateľských rozhraní, alebo poskytovanie rozhraní API na integráciu iných systémov. Vyhľadávacie služby sú zabezpečené prostredníctvom popisných metadát a textového obsahu [72].

Skutočnosť, že prijaté dáta zostanú autentické sa zaisťuje aj zaznamenávaním metadát PREMIS zakaždým, keď sa vykoná akcia na digitálnom objekte. Všetky interakcie medzi používateľmi a archívom sa zaznamenávajú z bezpečnostných dôvodov ako aj z dôvodu zodpovednosti za vykonané akcie. RODA umožňuje pridanie

ďalších funkcionalít pomocou rozšírení, ako napríklad hodnotenie platnosti digitálnych podpisov počas prijímania, ako aj schopnosť znovu podpísať archivované dokumenty po skončení platnosti ich podpisu [72], [73].

Široká škála základných ako aj doplňujúcich funkcionalít zabezpečuje so skutočnosťou otvoreného zdrojového kódu dobrý predpoklad na implementáciu vyhovujúcu rôznym potrebám. Existuje aktívna komunita, ktorá spolu s vývojármi neustále zlepšuje a inovuje vlastnosti tohto archivačného systému.

6.4 ESSArch

ESSArch je archívne riešenie s otvoreným zdrojovým kódom, ktoré je založené na modeli OAIS. Bolo vyvinuté v spolupráci s viacerými európskymi archívmi, a neskôr sa stalo súčasťou projektu E-ARK, kde získalo hodnotenie „vynikajúci“, čo bolo najvyššie možné hodnotenie. Systém je súčasťou aj projektov (EARK4ALL, E-ARK3), preto sa neustále vyvíja a všetky nové štandardy sú do neho ihneď implementované. Podporuje tvorbu a správu rôznych formátov metadát (METS, PREMIS) [73].

ESSArch pozostáva zo softvérových komponentov, ktoré poskytujú komplexné funkcie pre archiváciu ako je predbežný príjem, príjem, správa dát, archívne úložisko, správa archívu, a ďalšie. Každú časť ESSArch je možné používať jednotlivo, ako aj ľahko integrovať, aby bola zaistená celková funkčnosť pre producentov dát, archivárov a spotrebiteľov. Vývojári zabezpečili podporu pre nasadenie v kontajnerovom prostredí. Dostupná dokumentácia je prehľadná, ale odráža len základné nastavenie a funkcionalitu systému. Neexistuje veľká komunita používateľov, alebo aktívne fórum, pretože systém je nasadený väčšinou vo veľkých archívoch rôznych inštitúcií [74].

6.5 E-ARK Web

E-ARK Web je archivačný systém s otvoreným zdrojovým kódom, ktorý využíva webové rozhranie pre komunikáciu s používateľom. Je zameraný na model OAIS, čo znamená, že systém má implementované všetky funkčné entity. Archivačné balíky vytvára podľa štandardu E-ARK. Podporuje širokú škálu metadát ako sú popisné metadáta EAD, štrukturálne METS či archivačné PREMIS. Tento archivačný systém bol vytvorený v európskom projekte E-ARK ako demonštrátor nových štandardov a archivačných postupov. Aj v súčasnosti participuje v projekte E-ARK3 a preto je neustále vyvíjaný. Cieľom vývoju je, aby výsledné riešenie bolo dostupné, ľahko implementovateľné bez potreby nasadenia špeciálnych podnikových riešení, alebo odbornej pomoci. Vývojári pripravili systém aj pre kontajnerové prostredie.

Keďže sa jedná stále o nástroj vo vývoji, dokumentácia nie je podrobná a taktiež neexistuje v aktuálnej dobe nejaká komunitná podpora [30] [76].

6.6 Porovnanie vybraných archivačných systémov

V nasledujúcej tabuľke 6.1 sú porovnané vybrané archivačné systémy. Porovnanie je rozdelené na jednotlivé kategórie od všeobecných vlastností cez podporované štandardy až po jednotlivé časti modelu funkčnosti OAIS. Ohodnotenie je typu **áno** (obsahuje/spĺňa), **nie** (neobsahuje/nespĺňa), **áno/nie** (čiastočne obsahuje/spĺňa, ale je potreba zlepšenie, ďalší vývoj).

Tab. 6.1: Porovnanie vybraných archivačných systémov.

	Archive-matica	**DAITSS**	**RODA**	**ESSArch**	**E-ARK Web**
Všeobecné vlastnosti					
Otvorený zdrojový kód	áno	áno	áno	áno	áno
Webové rozhranie	áno	áno	áno	áno	áno
Virtualizácia	áno	áno	áno	áno	áno
Kontajnerizácia	áno	áno/nie	áno	áno	áno
Dokumentácia	áno	áno/nie	áno	áno	áno/nie
Ľahko rozšíriteľné	áno	nie	áno/nie	áno/nie	nie
Komunitná podpora	áno	nie	áno	nie	nie
Aktívny vývoj	áno	áno/nie	áno	áno	áno
Podporované štandardy					
OAIS	áno	áno	áno	áno	áno
METS	áno	áno	áno	áno	áno
PREMIS	áno	áno	áno	áno	áno
Dublin Core	áno	nie	áno	áno	áno
Špeciálne dátové spojenie (Data Submission Session)					
Kontrolný súčet	áno	áno	áno	áno	áno
Kontrola vírusov pred príjmom dát	áno	áno	áno	nie	nie
Dohoda o podaní s kontrolou integrity	áno	áno	áno	áno	áno
Príjem					
Kontrolný súčet	nie	áno	áno	áno	áno
Kontrola vírusov po príjme dát	nie	áno	áno	nie	nie

Archívne úložisko					
Pravidelný kontrolný súčet	nie	áno	áno	áno	áno
Pravidelná kontrola vírusov	nie	áno	áno	nie	nie
Plánovanie uchovania dát					
Pravidelná kontrolaformátov, prípadná migrácia	áno/nie	nie	áno	áno/nie	áno/nie
Správa archívu					
Správa dát	áno	áno	áno	áno	áno
Správa prístupu	áno	áno	áno	áno	áno
Zaznamenávanie aktivity v systéme	áno	áno	áno	áno/nie	áno/nie
Sprístupnenie					
Vyhľadávanie obsahu	áno	áno	áno	áno	áno
Vyhľadávanie metadát	áno	áno	áno	áno	áno
Obsah dostupný cez webové rozhranie	áno	áno	áno	áno	áno
Obsah dostupný priamo (bez špeciálneho zobrazovača)	áno	áno	áno	áno	áno

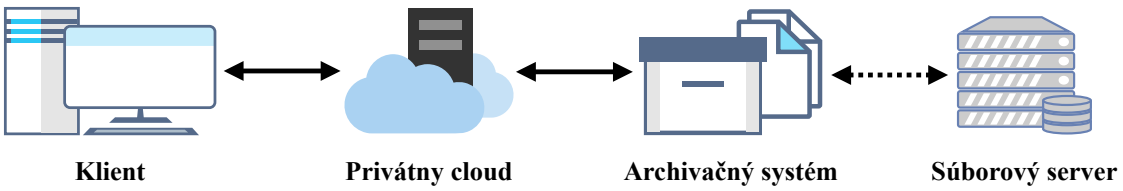
6.6.1 Zhodnotenie analýzy vybraných archivačných systémov

Archivačný systém pre dlhodobé uchovávanie dát by mal spĺňať široko rozšírené normy akou je napríklad OAIS, pretože tým sa definujú základné procesy nevyhnutné k dlhodobej ochrane dát. Potrebná je aj rozsiahla podpora metadátových formátov (METS, PREMIS, Dublin Core) pre zabezpečenie dlhodobého uchovávania. Dôležitou súčasťou je aj aktivita vývojárov a komunity pre chod tohto systému v čo najväčšom časovom období. Preto sa z tabuľky vybraných archivačných systémov javí ako vhodné riešenie Archivematica, ktoré spĺňajú hlavné požiadavky na archivačný systém. Archivematica je jedným z najviac rozšírených produktov, má širokú ponuku doplňujúcich funkcií a zároveň podporuje integráciu s viacerými systémami tretích strán.

7 Návrh privátneho cloudu s podporou dlhodobej archivácie

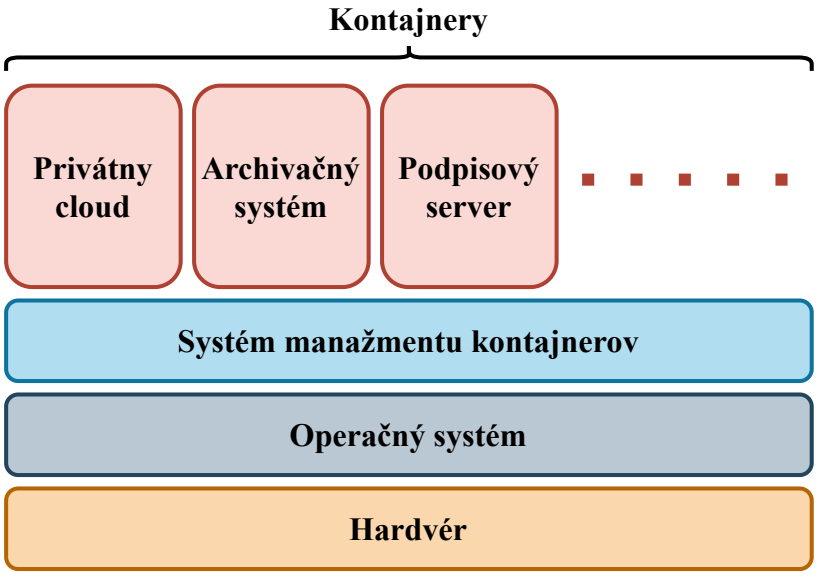
Hlavnou myšlienkou návrhu, je transformácia techník dlhodobej archivácie, ktoré sa využívajú vo veľkých inštitúciách a ich repozitároch do podoby použitia lokálneho rázu. To znamená, že aj malé firmy či jednotlivci budú schopní uchovávať údaje v rovnakej podobe za dodržania prísnych štandardov. Väčšina archivačných riešení je ale stavaná na podnikovej úrovni, kedy náklady za licencie, podporu a prevádzku bývajú nemalé. Preto návrh počíta s využitím systémov s otvoreným zdrojovým kódom, ktoré sú poskytované v nejakej verzií bezplatne. Taktiež formuje celý proces archivácie do čo najkompaktnejšej podoby, aby bolo možné archivačný systém prepojiť s privátnym cloudom pre uchovávanie dát a nasadiť ako celok aj na nie moc výkonné zariadenia. Myšlienkou je jednoduchosť obsluhy zo strany používateľa a maximálne šetrenie výpočtového výkonu bez zníženia úrovne bezpečnosti.

Základom návrhu je privátny cloud pre potreby ukladania dát, ktorému asistuje archivačný systém zabezpečujúci proces prípravy dát na dlhodobú archiváciu z tohto cloudu, viď obr. 7.1. Oba systémy musia byť prístupné z webového rozhrania, aby bola pre používateľa zabezpečená multiplatformovosť a jednoduchosť obsluhy. Archivačný systém bude môcť ukladať výsledné archivované súbory späť na cloud do oddelenej sekcie alebo na samostatný súborový server, ktorý môže podporovať prípadné prepojenie s verejným cloudom pre potreby záloh alebo migrácie archivovaných dát v budúcnosti.



Obr. 7.1: Základný zjednodušený návrh riešenia.

Návrh predpokladá implementáciu na jeden fyzický server, preto kvôli šetreniu výpočtového výkonu bola zvolená odľahčená virtualizácia prostredníctvom izolácií jednotlivých systémov do kontajnerov. Tieto kontajnery bežia na jednom spoločnom operačnom systéme, viď obr. 7.2, a preto oproti klasickej virtualizácií dokážu flexibilnejšie využiť výpočtový výkon. Tým pádom je možné jednoducho pridať do riešenia ďalšie podporné systémy podľa potreby. Základnom návrhu je teda operačný systém, ktorý by mal zabezpečiť dôvernosť obsahu ako aj vytvorenie redundancie dát pre zabezpečenie proti strate pri chybe disku (RAID). Taktiež by mal byť schopný vytvoriť a spravovať kontajnerové prostredie za pomoci manažéra kontajnerov Docker.



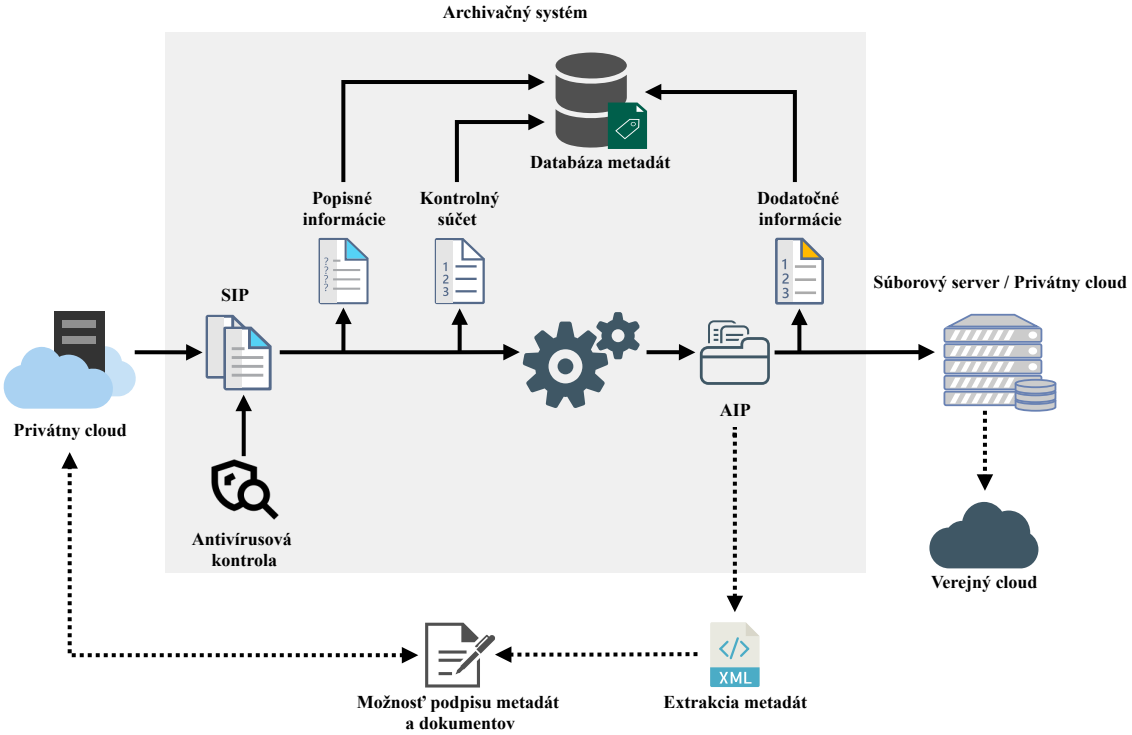
Obr. 7.2: Návrh implementácie pomocou kontajnerov.

Privátny cloud musí pozostávať zo softvéru, ktorý je stavaný na správu a ukladanie dát od používateľov cez prehľadné webové rozhranie zabezpečujúce jednoduchosť obsluhy a multiplatformovosť na strane klienta. Softvér musí byť kompatibilný s nasadením do kontajnerového prostredia. Keďže ide o vstupnú bránu pre bežných používateľov do celého riešenia, je nutné aby zabezpečoval dôkladnú správu prístupu do systému ako aj k jednotlivým súborom. Všetky vstupné dáta by mali prejsť antivírusovou kontrolou. Z pohľadu správcu musia byť k dispozícií prehľadné a kompletné logy zaznamenávajúce chod systému ako aj činnosť jednotlivých používateľov. Samozrejmosťou je taktiež otvorený zdrojový kód a dostupná podrobná dokumentácia.

Archivačný systém je navrhnutý ako celok, ktorý bude vykonávať prenos a spracovanie dát do archívu, vytváranie a uchovávanie potrebných metadát ako aj správu uložených archívov. Návrh podlieha medzinárodnému referenčnému modelu OAIS (ISO 14721) pre archivačné systémy. Posúdenie archivačného systému z pohľadu organizácie NDSA musí byť na úrovni 1, viď tab. 3.2. Archivačný systém by mal byť naviazaný na privátny cloud z ktorého môžu užívatelia vybrať dáta, ktoré je potrebné dlhodobo archivovať. Prístup do archívu je navrhnutý taktiež za pomoci prehľadného webového rozhrania a na základe povolenia podľa prístupových práv. Podporovať by mal spracovanie čo najviac štandardizovaných typov súborových formátov, aby bola docielená komplexná archivácia všetkých potrebných dát.

Samotný archivačný proces začína výberom potrebných dát používateľom na privátnom cloude, ktoré presunie do určeného umiestnenia. Následne v archivačnom systéme potvrdí výber dát, ku ktorým doplní potrebné metadáta, ktoré sa priložia ako popisné informácie k vybraným súborom. Systém vytvorí kontrolné súčty

jednotlivých položiek a ako celok sa všetky údaje spracujú do formy SIP balíka podľa vybraného štandardu. Pri tomto procese prebehne antivírusová kontrola všetkých vstupných údajov. Následne sa zo SIP generuje AIP (a príslušné metadáta k tomuto procesu), ktorý sa posiela na dlhodobú archiváciu na vybrané úložisko (primárne privátny cloud). Popisné informácie, kontrolný súčet a metadáta o generovaní AIP a jeho umiestnení, sa uložia do databázy, kde spolu predstavujú informácie zabezpečujúce integritu, autenticitu ako aj časové ukotvenie dát a jednotlivých krokov systému na základe požiadavky používateľa. Archivačný systém musí tieto informácie (metadáta), vedieť generovať podľa štandardov PREMIS a METS. Výsledný súbor metadát by sa mal dať vyextrahovať zo systému pre potreby nezávislej validácie dát z AIP. Celý proces archivácie by mal byť prehľadný s možnosťou čiastočnej alebo úplnej automatizácie.



Obr. 7.3: Návrh procesu archivácie.

Návrh počíta aj s možnosťou, že na niektoré dokumenty pred archiváciou bude kladený dôraz na zabezpečenie autenticity a nepopierateľnosti. Pre tieto prípady by mala byť možnosť priamo z privátneho cloudu vybrané dokumenty opatriť digitálnym podpisom buď externe alebo interne na vlastnom podpisovom serveri. Taktiež sa môžu podpísať vyextrahované metadáta z archivačného systému, viď obr. 7.3.

Z pohľadu bezpečnosti musia všetky komponenty návrhu používať iba publikované a overené kryptografické algoritmy a protokoly. Nesmú teda využívať prelomiteľné algoritmy (MD4, MD5, DES, RC4, SHA-1 atď.), či nie moc bezpečné módy

šifier (ECB, atď.). Dĺžka kľúčov bezpečnostných mechanizmov musí spĺňať minimálne požiadavky NIST a ECRYPT do roku 2030 respektíve 2028, viď tab. 4.1. V prípade archivačného systému je potrebné zvoliť najbezpečnejšie varianty z dostupných kryptografických algoritmov, pretože pri dlhodobej archivácií sa musí rátať so zastaraním aktuálneho zabezpečenia.

8 Možné formy implementácie návrhu

Táto kapitola popisuje viaceré možnosti implementácie na základe návrhu archivačného systému a vybraných existujúcich riešení. Približuje tiež funkcionalitu, náročnosť obsluhy ako aj iné vecné poznatky z testovania kompatibility a spoločného behu systémov. Taktiež hodnotí a porovnáva jednotlivé kombinácie z rôznych pohľadov.

Podstata každej popisovanej implementácie návrhu je rovnaká, rozdielom sú len použité archivačné systémy a ďalšie aplikácie, služby s nimi späté. Podľa návrhu na hardvéri beží operačný systém, v tomto prípade Rockstor. Ďalšia vrstva obsahuje manažér kontajnerizácie Docker. Základnými systémami, aplikáciami vo forme kontajnerov sú vždy privátny cloud (Nextcloud) a archivačný systém. Testovanie jednotlivých konceptov prebieha vo forme zostavenia a behu kontajnerov na serveri spoločne (stack) prostredníctvom jedného docker-compose súboru.

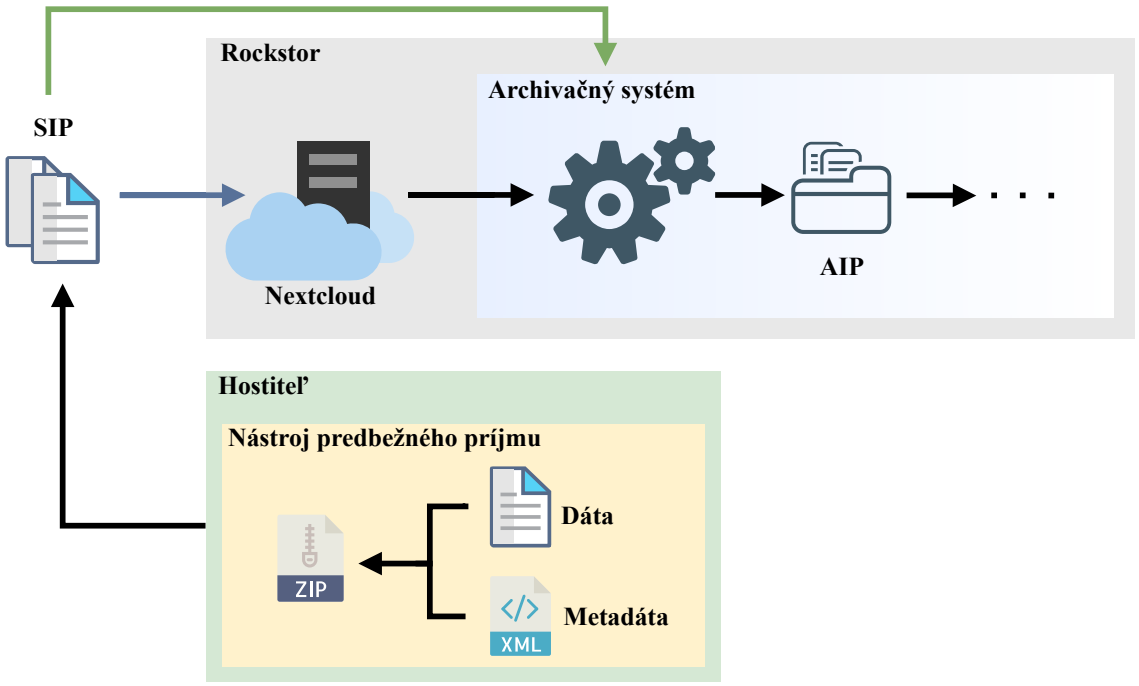
8.1 Rockstor

Rockstor je upravený operačný systém do podoby privátneho cloudu respektíve NAS riešenia s otvoreným zdrojovým kódom. Spája jednoduchosť a komplexnosť správy v cloude s dôkladným manažmentom fyzických diskov, ktorých kapacitu je možné ľahko rozšíriť alebo zúžiť pridaním alebo odstránením diskov priamo z webového používateľského rozhrania [77]. Ponúka funkcie ako miešanie diskov rôznych typov a veľkostí (pevné disky sa môžu kombinovať s externými (USB), atď.), pridávanie a odstraňovanie diskov, a mnohé ďalšie. Zabezpečenie uložených dát je umožnené šifrovaním celých diskov pomocou LUKS. Jadro systému tvorí buď operačný systém CentOS (staršie verzie) alebo OpenSUSE. Využíva súborový systém B-strom (Btrfs) a všetky jeho pokročilé funkcie. Vylepšenie základnej funkcionality je vyriešené netradične, a to pomocou takzvaných rozšírení „Rock-on“, čo sú vlastne kontajnerizované aplikácie, alebo služby (na báze Dockeru), ktorých existuje široká škála priamo podporovaných vývojármi, ale je možnosť si vytvoriť aj vlastné. Výhodou je taktiež plná funkcionalita bez potreby platenej verzie [78].

8.2 Cloudové riešenie so separovaným nástrojom na predbežný príjem

Vytváranie balíkov SIP prebieha používateľom manuálne, poloautomaticky alebo automaticky. Tento proces zabezpečujú nástroje predbežného príjmu. Nie vždy ale musia byť zabudované priamo do archivačného systému, a tak je potrebné ich nasadenie na zariadení používateľa. Postup archivácie je teda nasledovný. Vybrané dáta sa

zviažu s potrebnými metadátami na hostiteľskom zariadení používateľa a pretransformujú podľa potrebných štandardov nástrojom predbežného príjmu. Následne sa môžu balíky SIP priamo archivovať prostredníctvom archivačného systému, alebo sa uložia na cloudové úložisko a archivácia prebehne neskôr, alebo v presne definovaný čas, ale už bez potreby nahrávať dáta na server, viď obr. 8.1.



Obr. 8.1: Schéma riešenia so separovaným nástrojom na predbežný príjem.

Výhodou je, že balíky SIP môžu byť pripravené dopredu na ľubovoľnom zariadení a ich archivácia prebehne až pri možnosti prístupu na archivačný systém. Táto výhoda sa ale môže stať rýchlo nevýhodou, pretože síce balíky sú pripravené podľa štandardov, ale často mávajú samostatné archivačné systémy a ich repozitáre špeciálne požiadavky na hierarchiu a prípravu SIP balíčku. Prípadná chyba pri nastavovaní (vypĺňaní) tak vyjde najavo až pri spracovávaní na serveri, kedy sa oprava musí vykonať zase na lokálnom zariadení. Riešením je obstaranie takzvanej klasifikačnej schémy z archivačného systému, ktorá obsahuje všetky potrebné náležitosti pre SIP balíček. Ďalšou nevýhodou môže byť potreba inštalácie dodatočných aplikácii na strane používateľov.

8.2.1 Koncept 1

Tento Koncept spočíva vo využití aplikácií Exactly alebo RODA-in pre vytvorenie balíkov SIP lokálne na zariadeniach používateľov. Archivačný systém zastáva RODA vo verzii 4 a systém na správu a ukladanie súborov Nextcloud.

Príprava a nasadenie

Oba nástroje pre vytvorenie balíkov SIP je možné stiahnuť v podobe pre operačný systém MAC, Windows a ako Java súbor (JAR). Pri RODA-in nie je potrebná následná inštalácia, ale pre chod sa vyžaduje Java verzie 8.

Pre nasadenie Nextcloud a RODA na Rockstor vo forme Docker kontajnerov je potreba vytvoriť docker-compose súbor, kde sa treba zamerať hlavne na správne určenie portov a umiestnenia dát jednotlivých služieb (nastavenie volumes). Vývojári sami vytvárajú, udržujú a zverejňujú na Docker hub pripravené obrazy pre jednoduché použitie a nasadenie v kontajnerovom prostredí. V prípade RODA archivačného systému ide o obraz keeps/roda. Nextcloud sa použije v kombinácií s externou databázou (mariadb) v oddelenom kontajneri.

Súbor docker-compose tak obsahuje nastavenie 3 kontajnerov pre nasadenie spoločne (stack). V tejto zostave bez záťaže potrebujú približne **2,2 GB** pamäte RAM.

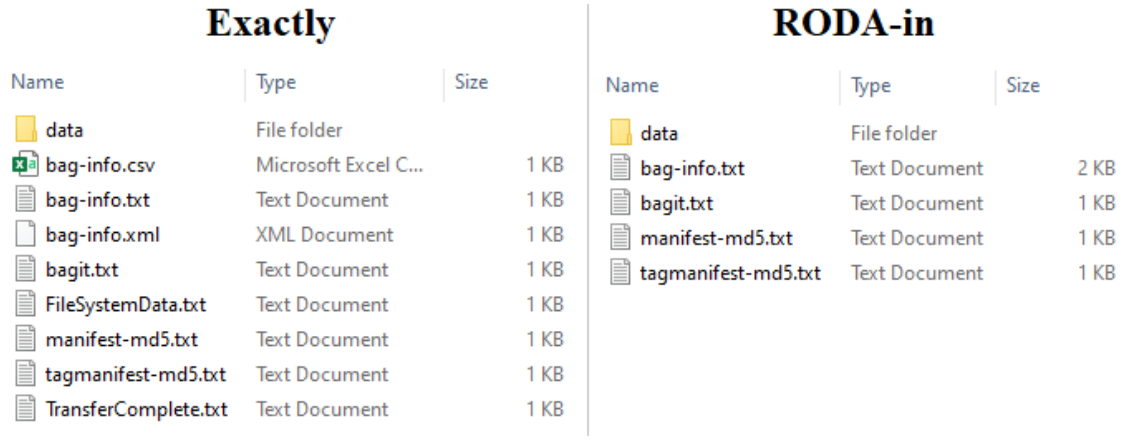
Upozornenie: po nasadení kontajnerov treba otestovať možnosť Nextcloudu zapisovať na pridané umiestnenia pre dáta mimo vyhradenej docker oblasti. Je bežné, že vo vnútri kontajneru sú zložky vytvárané a spravované iným používateľom (hlavne pri predpripravených obrazoch), ako mimo neho. Preto treba zabezpečiť zmenu na rovnakého vlastníka ku všetkému, k čomu má mať Nextcloud prístup. Taktiež sa stáva, že sa nestačí pripojiť do kontajnera a zistiť meno používateľa, ktorý je vlastníkom, pretože v hostiteľskom systéme sa eviduje zvyčajne iba pod UID. Preto sa odporúča rovno vyhľadať umiestnenie dát kontajner na hostiteľovi, zistiť identifikátor majiteľa, a ten nastaviť aj pre ostatné externé úložiská.

Základná funkcionalita

Aplikácia Exactly je prehľadná a obsluhuje sa jednoducho. Používateľ zadá meno výsledného SIP, umiestnenie zdroja a konečnú destináciu dát. Možné je aj vytvorenie jedného balíka z viacerých zdrojov (jednotlivé súbory alebo viaceré priečinky). Na výber má taktiež zabalenie do formátu Zip, ako aj priame zaslanie na server prostredníctvom FTP alebo SFTP. Ďalej treba zadať potrebné metadáta, kde program zobrazuje rezervované možnosti pre vyplnenie ako aj pridanie vlastných polí. Následne po stlačení tlačidla Transfer sa vytvorí BagIt balíček pripravený na archiváciu, viď obr. 8.2.

RODA-in obsahuje širokú škálu možností pre vytváranie SIP balíčkov. Používateľ si vyberie priečinok, ktorý chce pripraviť pre archiváciu. Tento nástroj umožňuje aj hromadné spracovanie dát, to znamená, že z veľkého počtu pridaných dát dokáže selektívne spracovať a pripraviť viacero informačných balíkov naraz podľa potrieb používateľa (asociácia). Na výber je možnosť selekcie súborov a priečinkov na spracovanie do jedného SIP, automatické vytváranie balíku pre každý súbor samostatne,

alebo balík pre všetky pridané dáta. Následne je potrebné vytvoriť metadáta, kde je možnosť výberu z viacerých predlôh (Dublin Core, EAD 2002, atď), ako aj import dopredu predpripraveného súboru. Po vyplnení všetkých metadát sa prejde na vytvorenie SIP balíku (Tlačidlo Create SIP), kde sa určujú základné nastavenia ako napríklad koncové umiestnenie, zadáva sa meno a výsledný formát, kde je na výber BagIt, E-ARK, E-ARK2. V tomto nástroji je taktiež možnosť importu klasifikačnej schémy vytvorenej archivačným systémom, pre zabezpečenie rovnakých požiadaviek na prípravu SIP balíka.

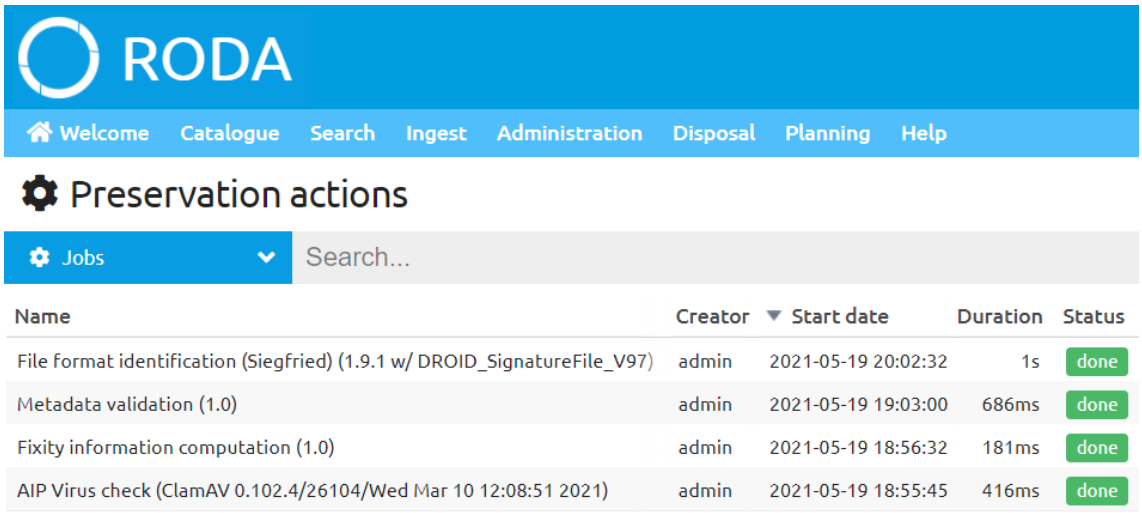


Obr. 8.2: Porovnanie základnej štruktúry BagIt balíka jednotlivých nástrojov.

Pre systém RODA sú základné prihlasovacie údaje sú `admin`, heslo `roda`. Predpripravené používateľské rozhranie je veľmi prehľadné. Pre archiváciu údajov stačí na karte Ingest v položke Transfer nahrať predpripravené SIP balíky. Následne v Ingest proces vytvoriť nový proces, označiť nahrané dáta a spustiť (start new process). Následne je na výber zo 4 pracovných tokov, ktoré sa dajú upravovať. Tie označujú, ktoré všetky úlohy sa vykonajú pri vytváraní balíka AIP. Na výber sú:

- Formát SIP balíka – používateľ musí vybrať podľa akého štandardu pre pripravený nahraný balík (BagIt, E-ARK, obyčajný priečinok alebo súbor brať ako SIP).
- Vybratie nadradeného uzla
- Antivírusová kontrola AIP - využíva ClamAV.
- Validácia metadát – táto možnosť je vždy aktívna.
- Vykonanie Fixity – kontrola obsahu AIP a vypočítanie kontrolných súčtov pomocou algoritmu SHA-256. Táto možnosť je vždy aktívna.
- Identifikácia formátu súborov – využíva nástroj Siegfried.
- Verifikovanie práv používateľa – táto možnosť je vždy aktívna.
- Automatické schválenie – nahrá súbor na úložisko (do repozitára).
- Emailová notifikácia po ukončení archivácie

Po tomto procese je výsledný balík AIP uložený v repozitári a dáta aj metadáta sú dostupné v karte Catalogue. Informácie a podrobnosti sú zobrazené veľmi prehľadne. Z pohľadu administrátora je možné dodatočne vykonať akcie pre celé úložisko, viď obr. 8.3, napríklad kontrola balíkov AIP antivírusovým systémom, alebo prepočítanie kontrolných súčtov. Logy zaznamenávajú všetky aktivity a sú dostupné aj štatistiky o procesoch a stave archivačného systému. Podrobná je aj správa prístupu, kedy používateľovi alebo skupine môže byť povolená či odmietnutá skoro každá možnosť v systéme.



Obr. 8.3: Dodatočné akcie pre repozitár vykonané administrátorom.

Hodnotenie konceptu 1

Koncept spočíva v jeho jednoduchosti a rýchlosti nasadenia. Pre základnú funkcionalitu postačujú predpripravené obrazy vývojármi, ktoré je možné doplniť a upraviť na mieru pre každé jedno nasadenie. Pre malý počet kontajnerov ide o riešenie s ľahkou správou a menšími nárokmi na výpočtový výkon. To ale len v prípade, že sa využíva v prostredí s nie veľkým počtom používateľov. Keďže ale tento koncept nevyužíva pre všetky kroky potrebné k archivácií cloudový výpočet, nespĺňa úplne zadanie práce. Tento variant bol testovaný hlavne pre demonštráciu iných možností nasadenia ako aj samotné nástroje predbežného príjmu pre vytvorenie balíka SIP, ktoré sa dajú použiť bez problémov v kombinácií s iným archivačnými systémami.

Nástroj Exactly je veľmi jednoduchý a prehľadný. Jeho výhodou je taktiež dobrá práca s metadátami, alebo možnosť automatického nahratia na server. Za nevýhodu možno považovať podporu tvorby výhradne balíkov BagIt, ako aj nemožnosť použitia predpripraveného súboru s metadátami, alebo klasifikačnú schému od archivačného systému.

RODA-in sa vyznačuje komplexnosťou, kedy podporuje viaceré štandardy SIP balíkov ako aj metadát. Výhodou je aj dobrá selekcia súborov medzi jednotlivé balíky. Podpora klasifikačných schém a šablón metadát uľahčujú prácu pri hromadnom spracovaní. Za nevýhodu možno považovať nemožnosť priameho zasielania balíkov do archivačného systému alebo na server. Široká škála možností sa prejavuje v neprehľadnosti, ako aj výskytu častých chýb, ktoré používateľa nútia reštartovať aplikáciu a tak začať celý proces prípravy nanovo.

Archivačný systém RODA obsahuje prehľadné predpripravené používateľské rozhranie, v ktorom sú všetky potrebné kroky vysvetlené a logicky oddelené. Chýbajú v ňom ale pokročilejšie nastavenia pre úložiská, proces spracovania dát a vytváranie AIP balíkov. RODA umožňuje prehľadne spravovať celý systém, vďaka podrobným logom s jednoduchým triedením, ktoré popisujú všetky aktivity v systéme. Taktiež sú dostupné štatistiky aj vo forme grafov. Veľkou nevýhodou je nemožnosť prípravy balíka priamo vo webovom rozhraní archivačného systému. Komunitná edícia dostupná zdarma má výrazne obmedzené možnosti, a podpora je len formou riešenia problémov na githube. Dostupná dokumentácia nie je úplná a často aj neaktualizovaná. Celkovo sa jedná ale o dobrú platformu, na ktorej sa dá vytvoriť prepracovaný systém archivácie, ako aj príslušný zdieľaný repozitár.

8.3 Kompletné cloudové riešenie

V nasledujúcich konceptoch sa celý proces archivácie ako aj príprava balíkov SIP vykonáva na serveri. Používateľ pracuje výhradne cez webové rozhranie, viď obr. 8.4, preto sa vyžaduje použitie moderného prehliadača.

8.3.1 Koncept 2

Tento Koncept spočíva vo využití archivačného systému ESSArch založenom na štandarde OAIS, podporujúci Európsku špecifikáciu E-ARK pre archiváciu digitálnych dát. Ako systém na správu súborov a uchovávanie balíkov AIP bol zvolený systém Nextcloud.

Príprava a nasadenie

ESSArch podporuje nasadenie v kontajnerovom prostredí. Vývojári predpripravili súbor docker-compose.yml, do ktorého treba doplniť systém Nextcloud, ktorý pre potreby databázy môže pre šetrenie výpočtového výkonu využívať kontajner archivačného systému Mariadb. Samotný ESSArch sa skladá z viacerých kontajnerov, kde väčšina ich obrazov je predpripravená, a tak sa len následne stiahnu z dockerhub:

- **Essarch** – tvorí jadro systému a je vytvorený zo zdrojového kódu lokálne pomocou Dockerfile.
- **Worker** – odvodený od essarch, vykonáva požiadavky na pozadí.
- **Beat** – odvodený od essarch, je potrebný na nepretržité prevádzkovanie niektorých operácií.
- **Elasticsearch** – slúži na vyhľadávanie, využíva predpripravený obraz verzie 7.6.1.
- **Rabbitmq** – zabezpečuje správu a distribúciu pracovných tokov v ESSArch, využíva predpripravený obraz (minimálna verzia 3.7.0).
- **Mariadb** – predstavuje databázu systému, využíva predpripravený obraz verzie 10.4.6.
- **Redis** – slúži ako vyrovnávacia pamäť na zlepšenie výkonu a na podporu upozornení v reálnom čase, využíva predpripravený obraz (minimálna verzia 3.0).
- **Logstash** – spolupracuje s elasticsearch a spracováva údajov z mnohých zdrojov, využíva predpripravený obraz.
- **Kibana** – používa sa na vizualizáciu elasticsearch dát, využíva predpripravený obraz.

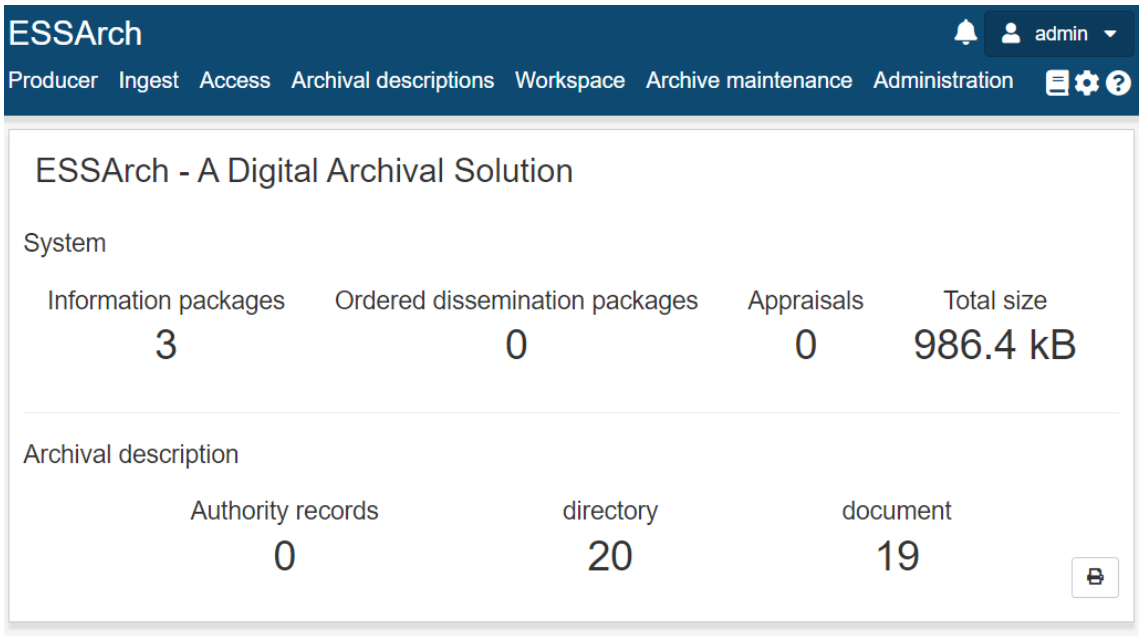
Pre nasadenie je potrebné najskôr naklonovať repozitár z githubu [79], kde sa v zložke docker nachádza docker-compose súbor. V ňom sa doplnia potrebné údaje pre Nextcloud a upravia porty, umiestnenia pre ukladanie dát jednotlivých kontajnerov. Elasticsearch vyžaduje zväčšenie maximálneho adresného priestor virtuálnej pamäte na 262144 (vm.max_map_count = 262144).

Súbor docker-compose tak obsahuje nastavenie 9 kontajnerov pre nasadenie spoločne (stack). V tejto zostave bez záťaže potrebujú približne **2,7 GB** pamäte RAM.

Základná funkcionalita Essarch

Základné prihlasovacie údaje do používateľského rozhrania sú `admin`, heslo `admin`. Na hlavnej obrazovke sú zobrazené údaje o počte archivovaných informačných balíkoch, ich celková veľkosť ako aj počet dokumentov a ďalšie údaje.

Pre archiváciu údajov je potrebné najskôr pripraviť informačný balík z ktorého sa vytvorí SIP. Tento proces prebieha v karte Producer. Balík sa pripraví podľa dohody o podaní (Submission Agreement), ktorá sa ustanovuje v sekcií settings. Následne v sekcií Collect content sa pridávajú dáta, ktoré je potrebné archivovať. Používateľ musí poznať štruktúru balíka daného štandardu, aby vedel kde presne, do akej zložky dáta nahrať, aby spracovanie prebehlo korektne. V kartách Create SIP a Submit SIP sa z pripravených dát vytvorí SIP, ktorý sa posunie procesu príjmu (Ingest), kde sa balík príjme a v sekcií Approval schváli pre archiváciu (vytvorenie AIP). Následne



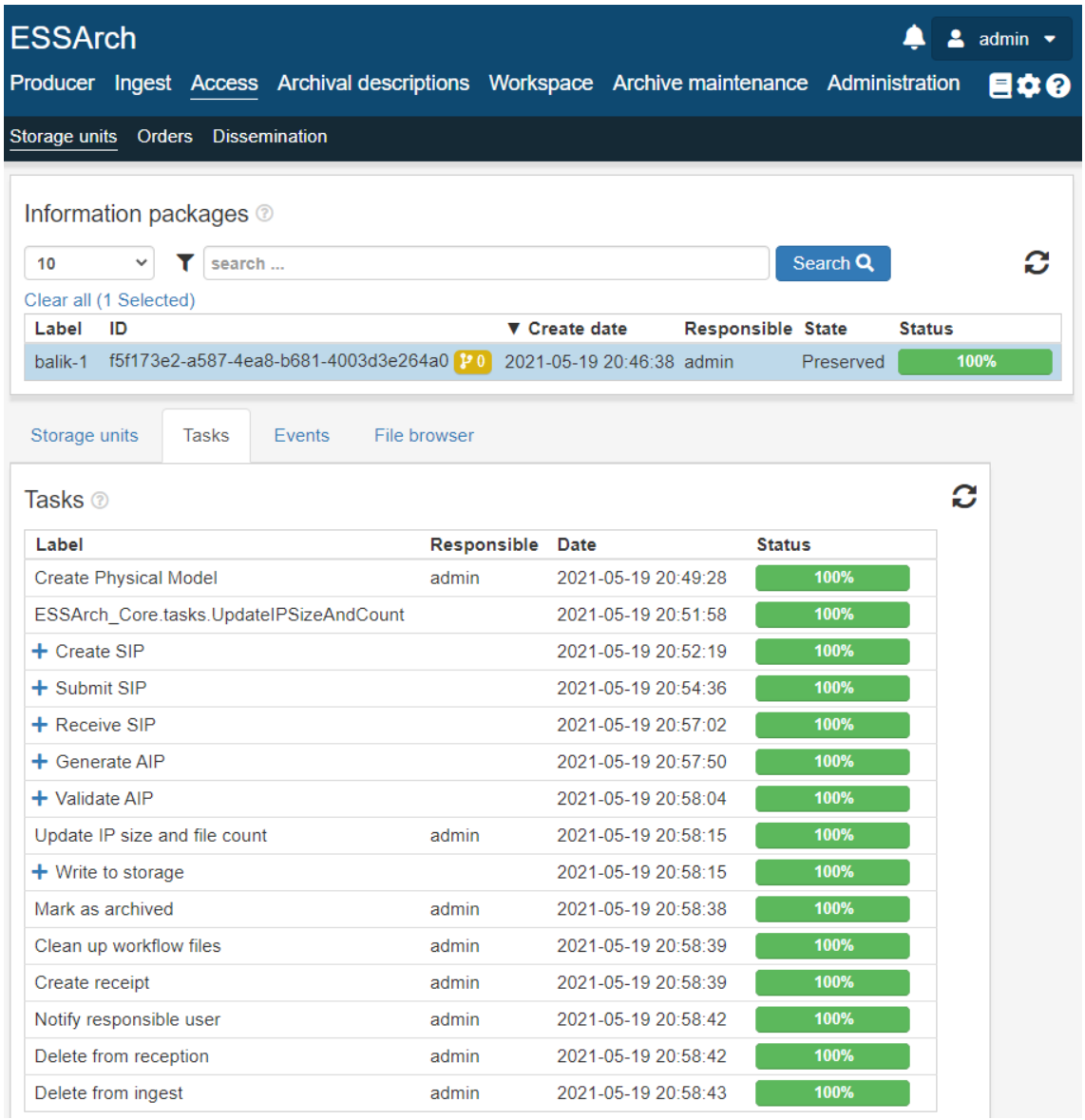
Obr. 8.4: Základné zobrazenie ESSArch so systémovými informáciami.

v sekcií Access používateľ vidí všetky archivačné balíky, kde si môže skontrolovať, či všetky potrebné úlohy boli vykonané korektne. Taktiež má prístup k obsahu balíka. Celý proces archivácie v ESSArch ponúka širokú škálu možností nastavenia balíkov, viď obr. 8.5, čo ho ale robí neprehľadný pre neskúseného používateľa, a preto treba rátať s nejakým časom na zácvik. Nevýhodou je aj to, že skoro všetky akcie treba potvrdzovať a nastavovať manuálne pri každom balíku a neexistuje jednoduchá možnosť aspoň čiastočnej automatizácie procesov.

Z administračného pohľadu sú k dispozícií podrobné informácie o jednotlivých úložiskách a ich obsahu (Sekcia Administration, karta Media information). V nastaveniach je možné zobrazenie a manažovanie všetkých informačných balíkov v systéme, ako aj pridanie dodatočných. Správca môže vytvárať jednotlivé dohody o podaní na základe požiadaviek repozitárov a taktiež jednotlivé profily pre štruktúry balíkov, alebo ich metadát, aby tak spĺňali požadované štandardy.

Hodnotenie konceptu 2

Koncept číslo 2 sa dá nasadiť do reálnej prevádzky bez väčších ťažkostí. Vývojári systém predpripravili aj pre použitie v kontajnerovom prostredí. Upravením docker-compose súboru sa dajú rôzne služby nastaviť či pridať. Samotný archivačný systém pozostáva s viacerých kontajnerov, čo viac zaťažuje systémové zdroje, ale dokáže tak lepšie zvládnuť spracovanie väčšieho množstva dát. ESSArch ponúka možnosť archivácie, od procesu výberu dát, až po vytvorenie a uloženie výsledného AIP



Obr. 8.5: Prehľad vykonaných úloh pre archiváciu balíka.

balíku. Používateľské webové rozhranie nie je moc prehľadné, a pre korektnú prácu so systémom treba dané osoby zaškoliť. Taktiež chýba možnosť automatizácie, kedy pre každý balík sa všetky úkony musia manuálne opakovať. Celkovo ide o stabilný systém, ktorý podporuje najnovšie formáty a štandardy pre archiváciu, akým je napríklad E-ARK.

8.3.2 Koncept 3

Tento koncept sa zakladá na použití archivačného systému E-ARK Web, ktorý predstavuje demonštrátor Európskej špecifikácie pre archiváciu E-ARK. Doplňujúci systém pre správu súborov a uchovávanie balíkov AIP je Nextcloud.

Príprava a nasadenie

Vývojári pripravili E-ARK Web pre nasadenie v kontajnerovom prostredí. Do predpripraveného docker-compose.yml súboru treba doplniť systém Nextcloud, ktorý môže využiť kontajner MySQL pre potreby databázy. Samotný systém E-ARK Web potrebuje pre svoj chod viacero kontajnerov:

- **Earkweb** – tvorí jadro systému a je vytvorený zo zdrojového kódu lokálne pomocou Dockerfile a podporných skriptov. Obsahuje viaceré služby ako napríklad Celery, ktorá sa stará o vykonávanie úloh prostredníctvom workerov (čím je ich viac tým rýchlejšie sa vykonajú požiadavky, ale je potrebný väčší výpočtový výkon) a Flower, ktorý monitoruje vykonávanie požiadaviek.
- **Solr** – slúži na indexovanie a vyhľadávanie obsahu. Využíva predpripravený obraz.
- **Redis** – využíva predpripravený obraz (minimálna verzia 3.4.1). Slúži ako vyrovnávacia pamäť.
- **RabbitMQ** – zabezpečuje sprostredkovanie správ medzi komponentami. Využíva predpripravený obraz.
- **MySQL** – zastupuje databázu systému. Využíva predpripravený obraz.

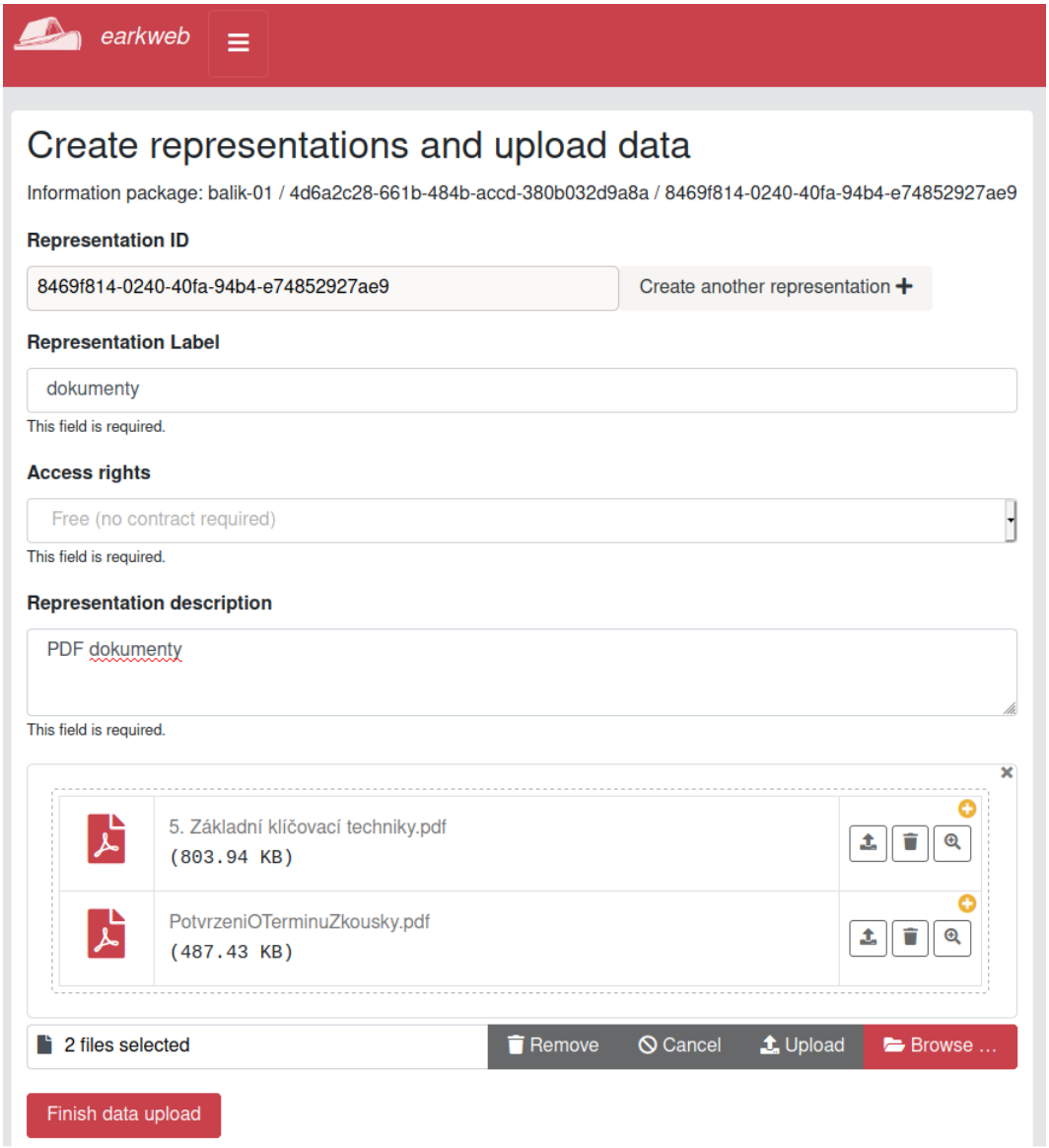
Súbor docker-compose tak obsahuje nastavenie pre 6 kontajnerov. V tejto zostave bez záťaže potrebujú približne **1,5 GB** pamäte RAM.

Pre nasadenie je potrebné najskôr naklonovať repozitár z githubu [80]. Keďže sa jedná stále o akýsi demonštrátor, je tento systém využívaný a upravovaný aj pre potreby aktuálneho vývoja v ďalších Európskych projektoch. Preto je možné, že v prípade nasadenia aktuálna verzia nemusí fungovať správne. Kvôli chýbajúcej podrobnej dokumentácií je prípadná oprava alebo ladenie systému časovo veľmi náročné a výsledok neistý. Pri testovaní bola najstabilnejšia verzia 1.0 z dátumu 31.7.2020 (`commit d7ce1ffcd2a1b0321235c2dcd4ee63f14308ee4c`), ale aj tá vykazovala občasné chyby.

Základná funkcionalita E-ARK Web

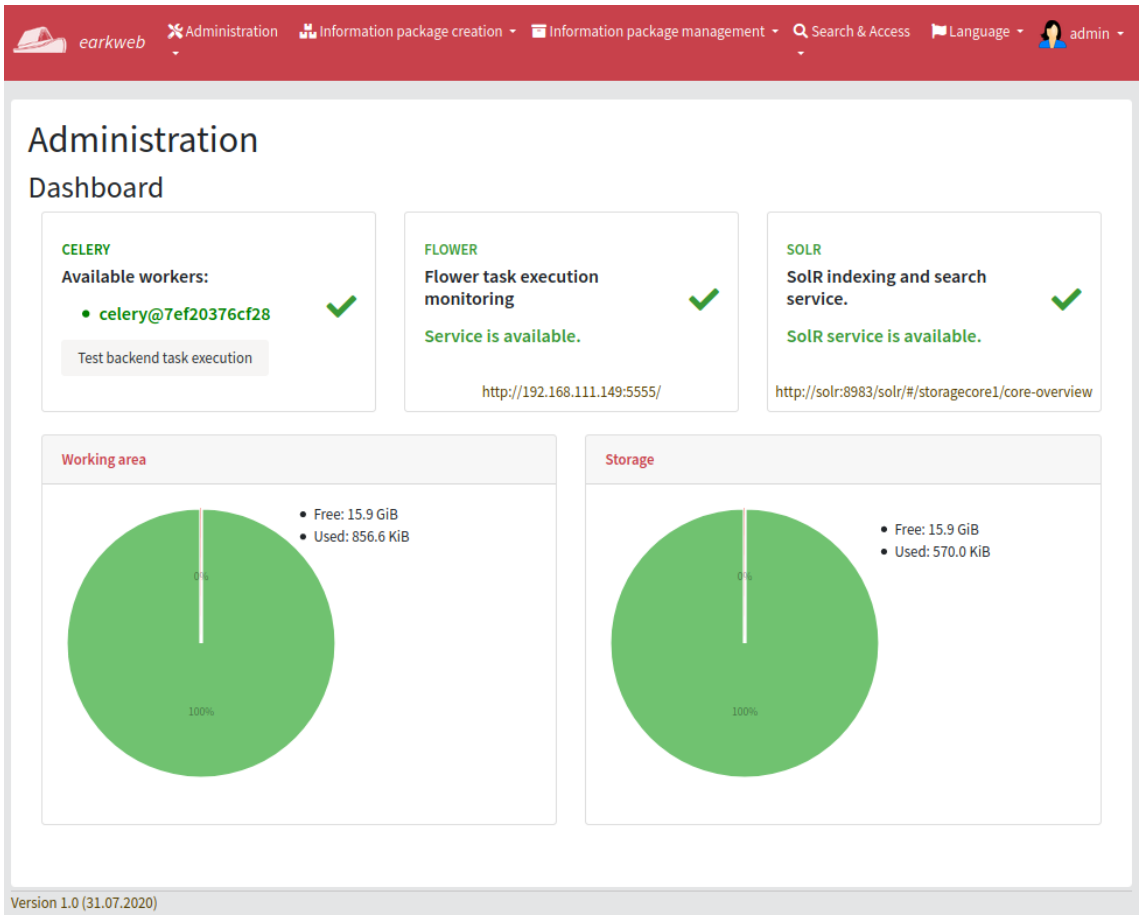
Základné prihlasovacie údaje do používateľského rozhrania sú `admin`, heslo `admin`. Webové rozhranie je prehľadné, a celá funkcionalita je rozdelená v hornom menu na administráciu, vytváranie balíkov, správu balíkov a vyhľadávanie. Pri vytváraní balíka sú všetky, kroky zoradené za sebou, a používateľ tak nemusí hľadať ako pokračovať v procese archivácie. V sekcii Information package creation (Create new information package) sa vytvára základ pre informačný balík. Po vyplnení všetkých potrebných metadát a pridelení reprezentačného ID sa pristupuje k nahrávaniu dát a nastaveniu prístupových práv, viď obr. 8.6. Po tomto kroku sa systém vytvorí balík SIP, ktorý je následne treba potvrdiť a rozhodnúť sa či ho ďalej archivovať

(tlačidlo Archive Information package). Systém využíva pre splnenie požiadaviek a formátu štandardu E-ARK funkcionalitu špeciálnych archivačných nástrojov (eArchiving Tool Box). E-ARK Web po procese archivácie umožňuje vyhľadávanie potrebných AIP v úložisku prostredníctvom názvu balíka, súboru, alebo len podľa typu prípony uložených dát. Používateľ nemá možnosť jednoduchej automatizácie procesov.



Obr. 8.6: Proces nahrávania dát do IP balíku v E-ARK Web.

Administrátor dokáže monitorovať všetky procesy pomocou Flower, a taktiež v sekcií Dashboard sú údaje o úložiskách a informácia o správnosti fungovania hlavných komponentov, viď obr. 8.7. Správa prístupu do systému sa nachádza v sekcií Django administration.



Obr. 8.7: E-ARK Web Dashboard.

Hodnotenie konceptu 3

Tento koncept sa neodporúča nasadzovať do reálnej prevádzky. Archivačný systém E-ARK Web je demonštrátorom aktuálnych projektov archivácie Európskej únie. Preto je tento systém v rozpracovanej fáze, kedy vykazuje mnohé nedokonalosti. Neexistujúca komplexná dokumentácia je tiež problémom hlavne pri výskyte nejakej chyby. Z používateľského hľadiska ide ale o prehľadné prevedenie procesu archivácie, kedy chýba len možnosť automatizácie. V budúcnosti po ukončení vývoja, môže ísť ale o výbornú a dostupnú platformu pre archiváciu podľa najnovších štandardov.

8.3.3 Koncept 4

Koncept číslo 4 zahŕňa použitie archivačného systému Archivematica, ktorá je jednou z najpoužívanejších platforiem pre archiváciu. Obsahuje vstavaného správcu dát a úložiska (Storage Service), ktorému v tomto koncepte asistuje systém Nextcloud.

Príprava a nasadenie

Archivematica podporuje nasadenie v kontajnerovom prostredí. Primárne je táto možnosť určená pre vývojárov, ale v poslednej dobe je považovaná za jeden z bežných typov nasadení. Predpripravený docker-compose.yml súbor je síce staršej verzie (2.1), ale aj tak sa dá bez problémov použiť a doplniť systém Nextcloud, ktorý môže využívať pre potreby databázy kontajner Archivematiky Mariadb. Samotný archivačný systém sa skladá z viacerých kontajnerov. Väčšina využíva predpripravený obraz stiahnuteľný z docker hubu: MySQL, Elasticsearch, Redis, Gearman, Fits, ClamAV, Nginx. Ostatné musia byť vytvorené lokálne zo zdrojového kódu: Archivematica MCP Server, Archivematica MCP Client, Archivematica Dashboard, Archivematica Storage Service.

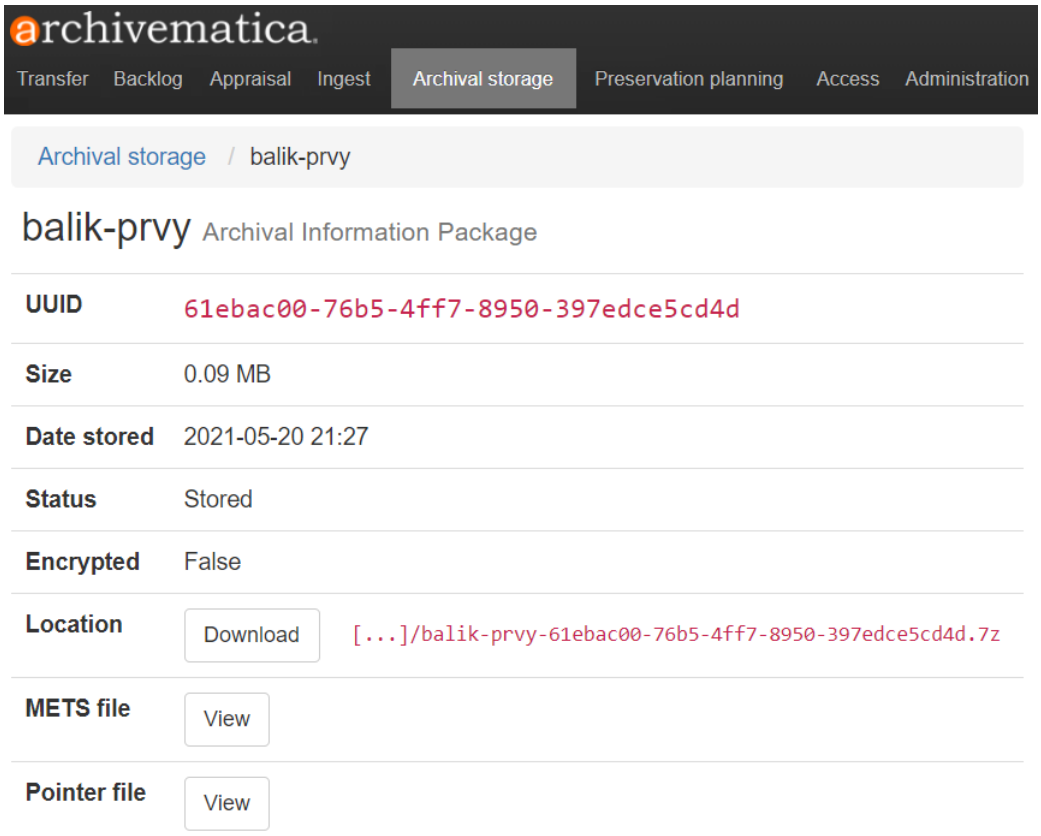
Pre vytvorenie kontajnerov je potrebné najskôr naklonovať repozitár z githubu, kde sa v zložke `am/compose` sa nachádza docker-compose súbor [81]. Do neho sa doplnia potrebné náležitosti pre systém Nextcloud. Elasticsearch vyžaduje zväčšenie maximálneho adresného priestor virtuálnej pamäte na 262144. Niekedy sa niektoré kontajnery nespustia, preto je nutný ich manuálny reštart.

Výsledný docker-compose tak obsahuje nastavenie 12 kontajnerov pre nasadenie spoločne ako stack. V tejto zostave bez záťaže potrebujú približne **2,6 GB** pamäte RAM.

Základná funkcionalita systému Archivematica

Základné prihlasovacie údaje do používateľského rozhrania sú `test`, heslo `test`. Dostupné webové rozhranie ponúka používateľovi menu s výberom možností podobným funkčným entitám modelu OAIS. V sekcií Transfer sa vyberajú dáta pre archiváciu ako aj pomenovanie budúceho balíka. Následne je potrebné pri každej akcii zvoliť jednu z možností. Ide o mikroslužby ako vytvorenie metadát, vypočítanie kontrolných súčtov či vírusový sken vstupných dát. Používateľ má tak možnosť dôkladne nastaviť presne to, čo chce aby sa s dátami pri archivácií dialo. Tento proces je ale zdĺhavý a pre neskúseného používateľa veľmi náročný. Preto Archivematika umožňuje správcom vytvoriť a nastaviť predlohy (Processing configuration), podľa ktorých následne systém automaticky spracuje a archivuje vybrané dáta. Používateľ tak len pri začatí procesu prenosu vyberie len potrebnú konfiguráciu. Podrobnosti o výslednom balíku AIP si je možné pozrieť v sekcií Archival storage, viď obr. 8.8.

Správu dát a úložiska zabezpečuje samostatná služba Storage Service, do ktorej sa treba samostatne prihlásiť. V nej sa vytvárajú a nastavujú presné lokácie pre výsledné AIP balíky, či miesta odkiaľ archivačný systém bude čerpať zdroje dát. Administrátor má široké možnosti správy, od manažovania používateľov, až po kontrolu úložísk či správnosti vykonania archivačných mikroslužieb.



Obr. 8.8: Zobrazenie podrobností AIP balíka v systéme Archivematica.

Hodnotenie konceptu 4

Koncept založený na systéme Archivematica, sa odporúča pre nasadenie v reálnom prostredí. Pri samotnej kontajnerizácií sa nevyskytujú problémy, ale príprava súboru docker-compose.yml môže pre jeho rozsah zabrať dosť času. Veľký počet kontajnerov sa neprejavil vo viditeľnom náraste potreby výpočtového výkonu v porovnaní s ostatnými podobnými archivačnými systémami. Z používateľského hľadiska ide o komplexné riešenie, ktoré umožňuje dôkladnú prípravu balíkov. Neskúsenej obsluhe pomáha v ovládaní možnosť automatizácie, ktorú dopredu pripravia správcovia pre jednotlivé typy dát. Z celkového hľadiska ide v aktuálnej dobe o výborné riešenie pre možnosti archivácie, ktoré je overené a stabilné. Dokumentácia je podrobná a často aktualizovaná. Taktiež pre veľkú používateľskú základňu sa prípadné problémy či nezrovnalosti dajú jednoduchšie riešiť.

8.4 Porovnanie a výber finálneho konceptu

Jednotlivé koncepty boli ohodnotené počas testovania podľa kritérií obsiahnutých v nasledujúcej tabuľke 8.1 (väčšia hodnota je lepšia). Najviac bodov získal kon-

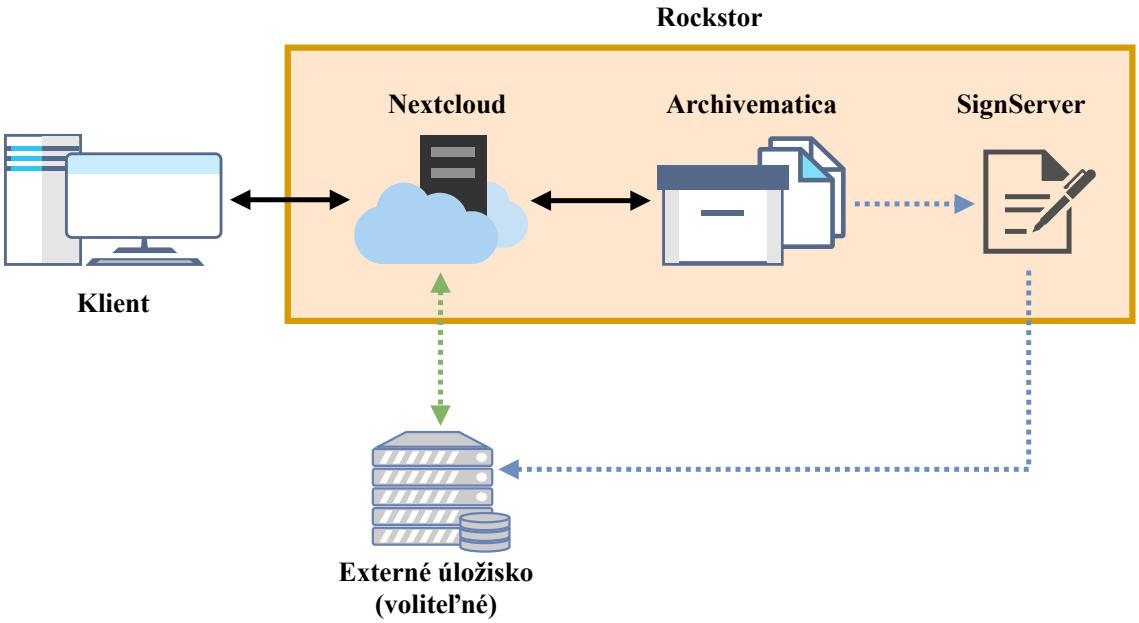
cept číslo 4, s archivačným systémom Archivematica, preto je vybraný do riešenia podľa návrhu privátneho cloudu s podporou dlhodobej archivácie.

Tab. 8.1: Hodnotenie konceptov v jednotlivých kritériách.

Koncept	**1**	**2**	**3**	**4**
Základná funkcionalita	6/10	8/10	7/10	10/10
Náročnosť obsluhy	8/10	6/10	9/10	7/10
Možnosti správy	7/10	8/10	6/10	10/10
Časová náročnosť implementácie	10/10	8/10	5/10	7/10
Hardvérové nároky	8/10	6/10	10/10	7/10
Celkový počet bodov	39	36	37	**41**

9 Implementácia konečného riešenia

Toto riešenie predpokladá nasadenie na jednom zariadení (serveri), viď obr. 9.1. Testovanie prebiehalo na zariadení s virtuálnym strojom, ktorému boli pridelené 2 fyzické jadrá procesora `Intel Core i5-4670K` (takt jadier 4,5 GHz), a 8 GB pamäte RAM. Procesor podporuje využitie inštrukcií novej generácie AES-NI. Na virtuálnom stroji beží systém Rockstor, ktorý v reálnych podmienkach bude nasadený priamo na hardvéri. Kontajnerizáciu zabezpečuje práve Rockstor. Táto kapitola popisuje základné nastavenie ako aj stručný návod na zavedenia a obsluhu jednotlivých komponentov. Taktiež upozorňuje v poznámkach na vyvarovaniu sa častých chýb pri konfigurácií. Zaznamenané sú aj hardvérové nároky ako aj zhodnotenie bezpečnosti dát. Celé konečné riešenie vychádza z návrhu privátneho cloudu s podporou dlhodobej archivácie 7, ako aj z kapitoly rozoberajúcej možné formy implementácie 8.



Obr. 9.1: Schéma konečného riešenia.

9.1 Základná konfigurácia Rockstoru

Vybrané riešenie je založené na systéme Rockstor. Ten je dostupný ako stabilná verzia 3.x.x, ktorej jadro tvorí Centos 7. Posledný predpripravený ISO obraz je 3.9.1 z roku 2017, preto sa neodporúča ho používať. Je tak potrebné vytvorenia vlastnej inštalácie za pomoci oficiálneho návodu z poslednej verzie 3.9.2 z roku 2020. Vývojári tento rok oznámili koniec ďalšieho vývoja pre Centos 7. Nástupníkom sa má stať

nová verzia 4.x.x, ktorá je založená na OpenSUSE Leap 15.2. Pre problémy s najnovšími funkciami Docker engine na Centos 7 a faktom, že všetky nové rozšírenia a podpora bude už iba v aktualizáciách pre OpenSUSE, je aj v tejto implementácií použitá zatiaľ vývojárska (testovacia) verzia, presnejšie 4.0.4. Ustanovenie a doladenie oficiálnej stabilnej verzie 4 je už skoro na konci, a dá sa očakávať jej zverejnenie ešte v tomto roku.

Celý proces prípravy systému nie je zložitý, ale keďže sa jedná stále o vývoj, oficiálna dokumentácia neexistuje, ale dostačujúca je na github a komunitnom fóre. Nová verzia môže byť zostavená na OpenSUSE Leap 15.2 a 15.3, ale prvý menovaný systém je doporučený a odskúšaný. Na zostavenie nového obrazu systému sa využíva nástroj kiwi-ng.

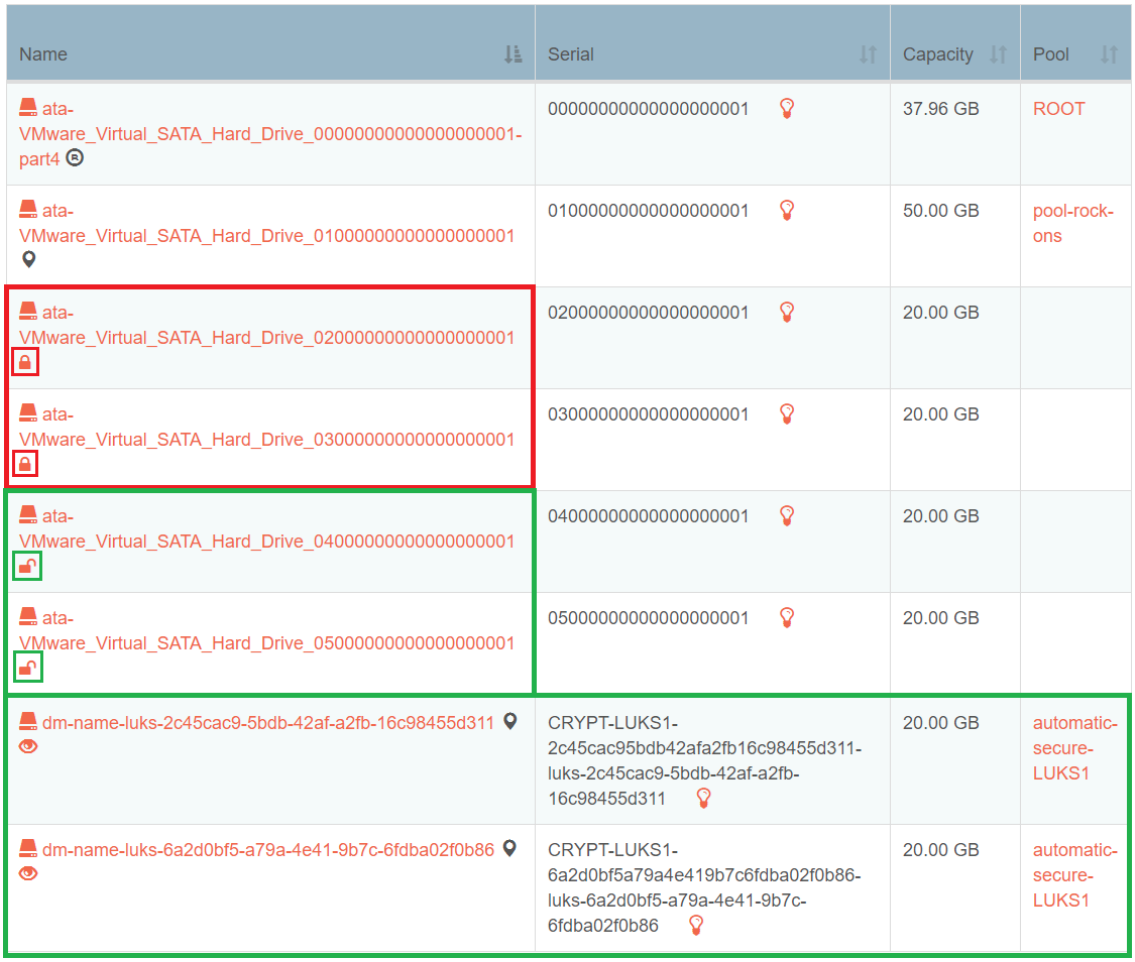
Pri následnom zavádzaní obrazu sa postupuje už podľa oficiálnej dokumentácie. To znamená nastavenie inštalačného disku, výber jazyka, zadanie hesla pre root používateľa. Po tomto procese a následnom prihlásení do konzoly je potrebné zadať príkaz myip, ktorý zobrazí pridelenú IP adresu pre webové rozhranie servera. V ňom je potrebné vytvorenie nového administrátorského účtu.

Základná konfigurácia sa dá zvládnuť časovo v pár desiatkach minút, výsledná veľkosť inštalačného iso súboru je okolo 500 MB.

Upozornenie: vždy po prihlásení za administrátora vo webovom rozhraní sa zobrazí možnosť aktivácie automatických aktualizácií. Táto funkcia sa do fázy, kedy nebude vydaná stabilná verzia Rockstor 4, neodporúča používať, pretože automatický prechod na nové vývojárske verzie nie je extra dopredu ohlásený a často spôsobuje stratu dát na serveri.

9.2 Nastavenie LUKS

Keďže sa očakáva, že niektoré dáta bude treba uchovávať s dôrazom na kryptografickú bezpečnosť, vybrané disky sú v riešení šifrované podľa štandardu LUKS. Táto možnosť je integrovaná priamo vo webovom rozhraní systému a tak nie je potreba používať konzolu. Nastavenie LUKS treba vykonať ešte pred namapovaním diskov. Všetky kroky sú podrobne popísané v systéme ako aj v príslušnej dokumentácií. Pre šifrovanie je potrebné zadať bezpečnostnú frázu (14 znakov minimum), ktorá následne slúži pri procese dešifrovania, respektíve odomknutia LUKS kontajneru, ktoré sa môže konať manuálne, vždy keď je potreba (alebo pri spúšťaní systému), a automaticky vytvorením špeciálneho súboru s kľúčom na odomknutie. Rockstor ho generuje ako 2048 bajtový pomocou `/dev/urandom` a ukladá na systémový disk v tvare `keyfile-` , kde za pomlčkou nasleduje identifikátor LUKS kontajnera (uuid). Na konci tohto procesu je potrebný reštart Rockstoru, aby zmeny boli viditeľné.



Obr. 9.2: Prehľad diskov v Rockstor s manuálnym a automatickým LUKS.

V tomto riešení automatické odomknutie súborom s kľúčom pri spustení systému nie je odporúčané použiť. Respektíve rapídne nezvyšuje bezpečnosť, pretože kľúč sa ukladá na nešifrované systémové úložisko. To znamená, že ak útočník bude mať fyzický prístup k zariadeniu (diskom), dokáže bez problémov súbory s kľúčmi extrahovať a dáta dešifrovať, preto je bezpečnejšie použitie konfigurácie kedy treba zadať bezpečnostnú frázu pri spustení systému na odomknutie šifrovaných diskov. Táto možnosť sa ale nejaví ako moc bezpečná pri citlivých dátach, pretože síce zabezpečuje disk, pokiaľ útočník nevie frázu, ale ak sa pri spustení systému odomkne, bude dostupný v nešifrovanej podobe po celý beh.

Pre tento problém sú dáta v tomto riešení rozdelené na tie, ktoré treba len archivovať bez nároku na kryptografickú bezpečnosť (dáta budú šifrované, ale nie počas behu systému), a dáta, ktoré je potrebné nie len uchovať ale aj utajiť. Tie sú šifrované rovnako pomocou LUKS, ale odomykajú sa manuálne len vtedy, keď je to nevyhnutne potrebné, viď obr. 9.2. Pre tento krok je už ale nevyhnutné použitie konzoly, ktorá sa dá obsluhovať aj priamo vo webovom rozhraní Rockstoru

po aktivácií služby `Shell In a Box`. Rockstor v základe vytvára LUKS verzie 1 so šifrou `aes-xts-plain64`, veľkosť kľúča 256 bitov. Ak je potrebná zmena parametrov musí sa použiť priamo nástroj `cryptsetup`, ktorý je už súčasťou systému a tak nie je potrebná jeho dodatočná inštalácia. Keďže návrh predpokladá nasadenie v prostredí s malým počtom používateľov, požiadavky na častú archiváciu citlivých dát nie sú očakávané. Preto, pokiaľ to výkon zariadenia dovoľuje, je dobré maximalizovať kryptografické zabezpečenie.

Najskôr pomocou príkazu `fdisk -l` treba zistiť označenia pripojených diskov (väčšinou tvar `sdX`). Následne sa vybraný naformátuje na LUKS s potrebnými parametrami, viď tab. 9.1, nasledujúcim príkazom:

```
cryptsetup -type luks2 -cipher aes-xts-plain64 -hash sha512 -iter-time
5000 -key-size 512 -pbkdf argon2i -sector-size 512 luksFormat /dev/sdX
```

Následne je používateľ vyzvaný na zadanie frázy, ktorá musí spĺňať bezpečnostné kritéria (dĺžka, unikátnosť, atď.), inak sa daná akcia nepotvrdí a príkaz bude treba zadať znova. Takto vytvorený šifrovaný LUKS kontajner sa ihneď zobrazí aj vo webovom rozhraní Rockstoru medzi diskami. Pre jeho sprístupnenie slúži príkaz:

```
cryptsetup luksOpen /dev/sdX názov
```

kde namiesto názov bude označenie disku/partície po dešifrovaní. Zatýmto krokom treba ešte zadať príkaz `mount`, ktorý pripojí sprístupnený disk medzi ostatné do systému. Kontrola tohto kroku sa dá spraviť príkazom `df -H` , ktorý zobrazí všetky pripojené disky/partície.

Po vykonaní všetkých potrebných operácií je potreba dáta znova previesť do šifrovanej podoby. Najskôr je potreba odpojenia disku pomocou `unmount /názov` a následne jeho uzamknutie `cryptsetup luksClose názov`.

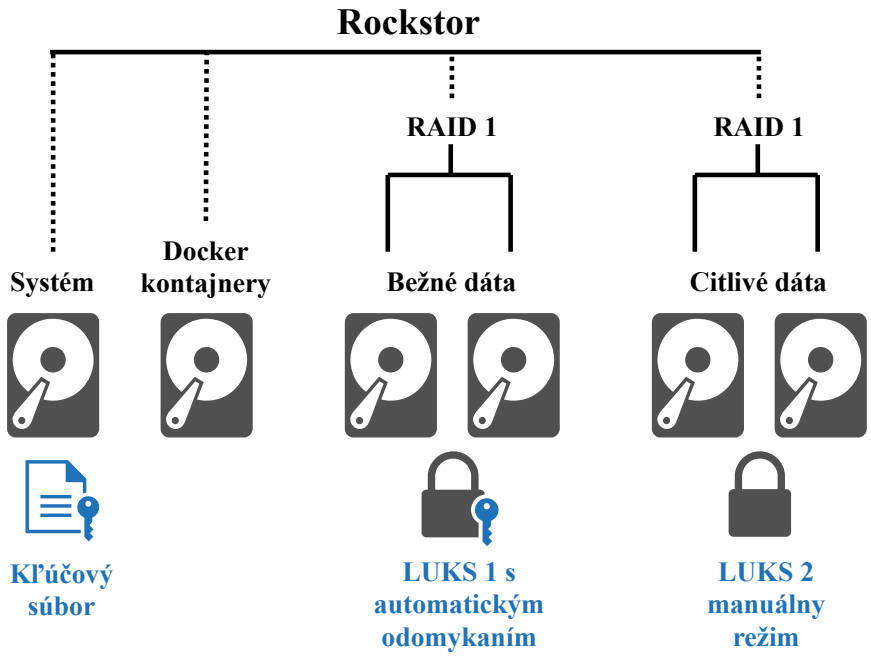
Tab. 9.1: Jednotlivé parametre zvolené pre celodiskové šifrovanie.

	Základné nastavenie Rockstorom	**Zvolené parametre pre citlivé dáta**
Verzia	LUKS 1	LUKS 2
Šifra	aes-xts-plain64	aes-xts-plain64
Veľkosť kľúča	256-bit	512-bit
Hashovací algoritmus	sha256	sha512
Odvodenie kľúča	PBKDF2	Argon2i
Iteračný čas (v milisekundách)	1000	5000

9.3 Konfigurácia RAID

Bezpečnosť dát pred náhlou stratou zabezpečuje použitie zrkadlenia v kombinácií so súborovým systémom Btrfs. Rockstor umožňuje rozdelenie a prácu so samostatnými partíciami (označované ako Pools), ale pre bezproblémový beh sa odporúča použitie celých diskov. V menu webového rozhrania pod Storage, v sekcii Pools sa vytvárajú partície, zrkadlenia a je tu aj možnosť kompresie. Na výber RAID 0, 1, 5, 6, 10, ale 5 a 6 sa neodporúča pre nekompatibilitu a chybovosť pri použití s Btrfs.

V základe je vždy vytvorený systémový pool ROOT. V tomto riešení sa odporúča použitie aspoň 6 samostatných diskov, viď obr. 9.3. Jeden pre systém, druhý pre Docker kontajnery a ostatné rozšírenia. Každý z nich musí mať samostatný pool. Pre bežné dáta je potreba vytvoriť pool z dvoch diskov, ktoré sú zabezpečené LUKS 1 a automatickým odomykaním pomocou kľúčového súboru, a nastaviť zrkadlenie pomocou RAID 1. Rovnaký postup sa uplatní aj pri citlivých dátach, len sa použijú disky s LUKS 2.



Obr. 9.3: Prehľad diskov v Rockstor s manuálnym a automatickým LUKS.

Upozornenie: všetky disky zabezpečené prostredníctvom LUKS, treba najskôr odomknúť a až následne bude umožnená možnosť vytvárania partícií (Poolov). Ak počas behu systému budú disky uzamknuté, vybraný Pool bude hlásiť stav nepripojený (unmounted), čo je v poriadku, pretože systém dokáže dáta čítať až po ich odomknutí. Netreba preto robiť žiadne dodatočné akcie.

9.4 Rock-ons

Rozšírenie funkcionality Rockstoru je možné pomocou Rock-onov, čo sú vlastne jednotlivé docker kontajnery. Túto možnosť treba najskôr aktivovať, a to v službách (services) zvolením Rock-on. Nutné je aj pridelenie úložného priestoru. Ten sa vytvára v sekcií Storage ako Shares (časť Poolov). Odporúčaná minimálna veľkosť pre tento účel je 5 GB. Ďalej je dobré pre každý novo pridaný Rock-on vytvoriť samostatný Share.

V menu v sekcií rock-ons je možné po stlačení tlačidla pre aktualizáciu (update) vidieť všetky natívne dostupné rozšírenia, ktoré sa dajú jednoducho nainštalovať (bežia samostatne ako kontajner). Celá funkcionalita je vlastne ovládanie častí Dockeru vo webovom rozhraní Rockstoru. Používateľ ale nie je obmedzený len možnosťou výberu s dostupných Rock-onov, ale dá sa vytvoriť aj vlastný.

Pridanie ďalších kontajnerov, ktoré sú potrebné napríklad pre archivačný systém a ďalšie služby by bolo touto cestou zdĺhavé a náročné, pretože webové rozhranie ponúka značne obmedzenú funkcionalitu a monitorovanie Dockeru. Možnosťou je aj správa priamo cez konzolu, čo je pri väčšom počte kontajnerov taktiež nie moc prehľadné a aj časovo náročné. Preto je odporúčané pridanie ako Rock-on rozšírenie ľubovoľný grafický nástroj na správu Dockeru. V tomto riešení zastupuje popisovanú funkcionalitu Portainer CE, ktorý je ale potrebé manuálne doplniť do Rockstoru.

Pridanie vlastného Rock-onu začína vytvorením json súboru s príslušným názvom (Portainer.json), viď výpis 9.1. Jeho obsah je veľmi podobný súborom docker-compose, ktoré popisujú vytvorenie kontajnera z príslušného obrazu a inštrukcií na jeho nastavenie. V tomto prípade je obraz dostupný na Docker hube (portainer/portainer-ce). Ďalej je potrebné nastaviť príslušný port, miesto pre ukladanie dát a rôzne popisné polia pre Rockstor. Takto pripravený súbor vložíme do špeciálneho adresára `/opt/rockstor/rockons-metastore`. Následne po aktualizácií v menu Rock-ons sa Portainer zobrazí ako možnosť na inštaláciu.

Upozornenie: adresár rockons-metastore nemusí existovať. V takom prípade ho treba vytvoriť a skontrolovať prístupové práva (malo by byť umožnené čítanie,zapisovanie aj spúšťanie).

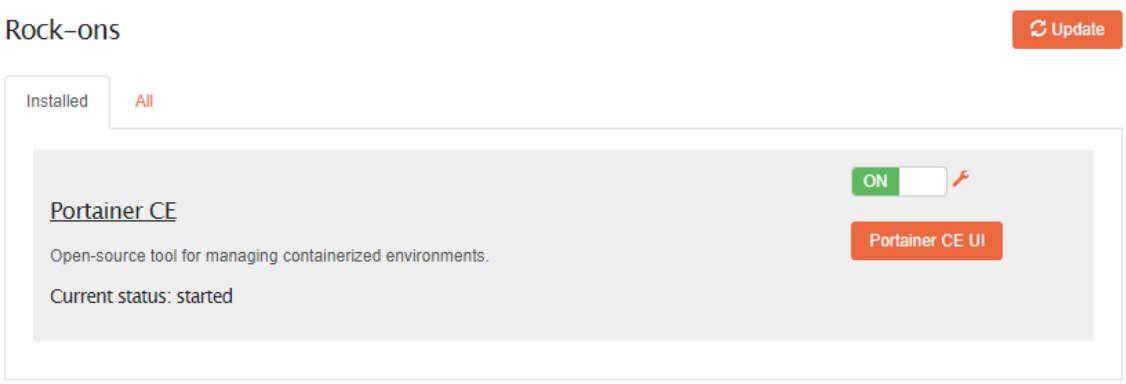
9.4.1 Nastavenie Portaineru

Po inštalácií rozšírenia, treba znovu obnoviť zoznam inštalovaných Rock-ons. Ak všetko prebehne v poriadku, aktuálny status bude ukazovať started, viď obr. viď obr. 9.4, a po kliknutí na tlačidlo používateľského rozhrania Portaineru (Portainer

Výpis 9.1: Portainer.json

```
{
  "Portainer CE": {
    "containers": {
      "portainer": {
        "image": "portainer/portainer-ce",
        "opts": [
          [
            "-v",
            "/var/run/docker.sock:/var/run/docker.sock"
          ]
        ],
        "launch_order": 1,
        "ports": {
          "9000": {
            "description": "Port for web ui",
            "host_default": 9000,
            "label": "Web UI",
            "ui": true
          }
        },
        "volumes": {
          "/data": {
            "description": "Persistent data storage",
              "label": "Data Storage"
          }
        }
      }
    },
    "description": "Tool for containerized environments.",
    "ui": {
      "https": false,
      "slug": ""
    },
    "website": "https://www.portainer.io/",
    "version": "1.0"
  }
}
```

CE UI), prebehne presmerovanie na príslušnú adresu a port v novom okne prehliadača. Ako prvé je potrebné vytvorenie administrátorského konta a následne vybratie koncového bodu (Endpoint), v tomto prípade ide o lokálny typu Docker.



Obr. 9.4: Portainer po pridaní ako Rock-on.

9.5 Príprava docker-compose

Predmetom návrhu je implementácia prostredníctvom docker kontajnerov spoločne (ako stack). Tým pádom je potreba vytvorenie súboru docker-compose.yml, v ktorom budú všetky potrebné náležitosti pre nastavenie a vytvorenie potrebných kontajnerov. Základom tohto súboru je archivačný systém Archivematica, pretože pre svoj chod potrebuje viaceré služby v samostatných kontajneroch a súborový server Nextcloud, ktorý je previazaný s úložiskom pre archiváciu. Ďalšie doplňujúce aplikácie ako podpisový server (Signserver) či certifikačná autorita (Ejbca) sa môžu pridať podľa potreby. Archivematica potrebuje pre svoj chod kontajnery so službami, kde niektoré majú dostupný pripravený obraz na stiahnutie, iné sa musia vytvoriť lokálne zo zdrojového kódu, preto je nutné na začiatku naklonovať repozitár z github [81], a taktiež stiahnuť všetky submoduly.

Samostatné služby Archivematiky (kontajnery):

- **MySQL** – zabezpečuje databázu pre chod archivačného systému, nasadenú ako Percona server. Využíva predpripravený obraz *"percona:5.6"*. Vyžaduje vytvorenie docker volume pre uchovanie dát.
- **Elasticsearch** – stará sa o indexáciu obsahu, a jeho následného vyhľadávania v procese archivácie alebo v samotnom úložisku balíkov. Využíva predpripravený obraz *"elasticsearch/elasticsearch:6.5.4"*. Vyžaduje vytvorenie docker volume pre uchovanie dát. Pre správny chod tohto kontajnera je potrebné zväčšiť maximálny adresný priestor virtuálnej pamäte. Túto zmenu je potrebné vykonať buď po každom spustení servera priamo príkazom `sysctl`

`-w vm.max_map_count=262144`, alebo uložiť danú hodnotu do súboru `/etc/sysctl.conf`.

- **Redis** – využíva sa na ukladanie a správu dočasných záznamov (cache) a sprostredkovanie správ. Využíva predpripravený obraz *"redis:3.2-alpine"*.
- **Gearman** – zabezpečuje riadenie fronty úloh, ktoré treba vykonať, aby bol proces čo najefektívnejší. Využíva predpripravený obraz *"artefactual/gearmand:1.1.17-alpine"*.
- **Fits** – tento kontajner obsahuje nástroj pre identifikáciu súborového formátu a jeho následnú validáciu. Využíva tak predpripravený obraz *"artefactual/fits-ngserver:0.8.4"*. Potrebuje mať zabezpečený prístup do všetkých umiestnení informačných balíkov.
- **ClamAV** – stará sa o antivírusovú kontrolu súborov. Využíva predpripravený obraz *"artefactual/clamav:latest"*. Taktiež potrebuje prístup do všetkých umiestnení informačných balíkov. Dôležité je aj nastavenie maximálnej veľkosti súborov na kontrolu, pretože veľké súbory môžu podstatne zvýšiť nároky na výpočtový výkon.
- **Nginx** – zastupuje webový server, na ktorom bežia služby webového rozhrania samotnej Archivematiky (dashboard) ako aj úložnej služby (storage service). Využíva predpripravený obraz *"nginx:stable-alpinet"*.
- **Archivematica MCP Server** – je jadrom archivačného systému. Riadi rôzne mikroslužby, a vykonáva požiadavky od používateľa zadané cez dashboard. Ďalej uchováva a vytvára logy o každej činnosti. Systém Archivematica sa pri spracovávaní príkazov spolieha na to, že klient a server majú prístup do rovnakých adresárov (základom je sharedDirectory, plus ďalšie voliteľné). Taktiež využíva služby MySQL a Gearman, preto musí byť s nimi prepojený. Obraz MCP Serveru musí byť vytvorený lokálne zo zdrojového kódu prostredníctvom Dockerfile (MCPServer.Dockerfile).
- **Archivematica MCP Client** – Archivematica môže mať jedného alebo viacero klientov, ktorý vykonávajú určité úlohy. Po spustení klient kontaktuje Gearman server, aké úlohy môže vykonávať, a čaká, kým mu nejakú nepridelí. Práve pre vykonávanie rôznych úloh musí byť prepojený s Fits, ClamAV, MySQL, Gearman, Elasticsearch a Archivematica storage service. Obraz MCP Clienta musí byť vytvorený lokálne zo zdrojového kódu prostredníctvom Dockerfile (MCPClient.Dockerfile).
- **Archivematica Dashboard** – je webové rozhranie, ktoré umožňuje používateľom spracovávať, monitorovať a riadiť procesy v archivačnom systéme. Potrebuje mať zabezpečený prístup k zdieľaným zložkám klienta aj servera, a tiež k službám MySQL, Gearman, Elasticsearch a Archivematica storage service. Obraz Dashboardu musí byť vytvorený lokálne zo zdrojového kódu

prostredníctvom Dockerfile (dashboard.Dockerfile).

- **Archivematica Storage Service** – je samostatná webová aplikácia, ktorá sprostredkováva presun súborov ku spracovaniu a následne do dlhodobého úložiska, kde udržuje informáciu o ich umiestnení pre neskoršie načítanie obsahu. Pozostáva z dvoch úrovní, kedy prvá hlavná Space označuje kde sú súbory uložené (napríklad lokálny súborový systém) a druhá Location zase určuje prečo sú tie dáta práve na tom mieste (napríklad zložka len pre AIP balíky). Potrebuje prístup k zdieľaným zložkám klienta aj servera a taktiež k umiestneniu pre zdroj dát na archiváciu. Využíva služby MySQL. Obraz Storage Service musí byť vytvorený lokálne zo zdrojového kódu.

Základné predpripravené nastavenie kontajnerov Archivematiky, sa nachádza v zložke compose v súbore docker-compose.yml. Do tohto konfiguračného súboru treba doplniť hlavne potrebné úložiská (volumes) pre balíky AIP, v tomto prípade ide o zrkadlené úložisko pre citlivé dáta šifrované LUKS 2 a taktiež zrkadlené úložisko pre bežné dáta šifrované LUKS 1 s funkciou automatického odomykania. Kontajnery majú k obom úložiskám prístup pre čítanie aj zápis (:rw).

```
- "./mnt2/data-secure-luks2/:/data-secure-luks2/:rw"
- "./mnt2/data-secure-luks1/:/data-secure-luks1/:rw"
```

Súborový systém nextcloud sa dá prostredníctvom kontajnerov implementovať rôznymi spôsobmi. V tomto riešení je použitý predpripravený obraz *"nextcloud"*, ktorý obsahuje vstavaný webový server Apache, obsluhujúci len potreby tejto aplikácie. Pre potreby databázy, bol kvôli optimalizácií zvolený variant využívajúci externé MySQL namiesto samostatného kontajnera, a to prostredníctvom už vytvorenej databázy pre Archivematicu, do ktorej sa stačí prihlásiť. Pre potreby uchovávania dát Nextcloudu je potrebné vytvoriť nové umiestnenie (volume), ako aj zabezpečiť prístup k už vytvoreným umiestneniam pre proces archivácie. Ide hlavne o priečinok pre zdroj dát archivačného systému pre proces vytvárania SIP (archivematica-sampledata), ako aj umiestnenia hotových AIP balíkov (sharedDirectory, data-secure-luks1, data-secure-luks2), viď výpis 9.2.

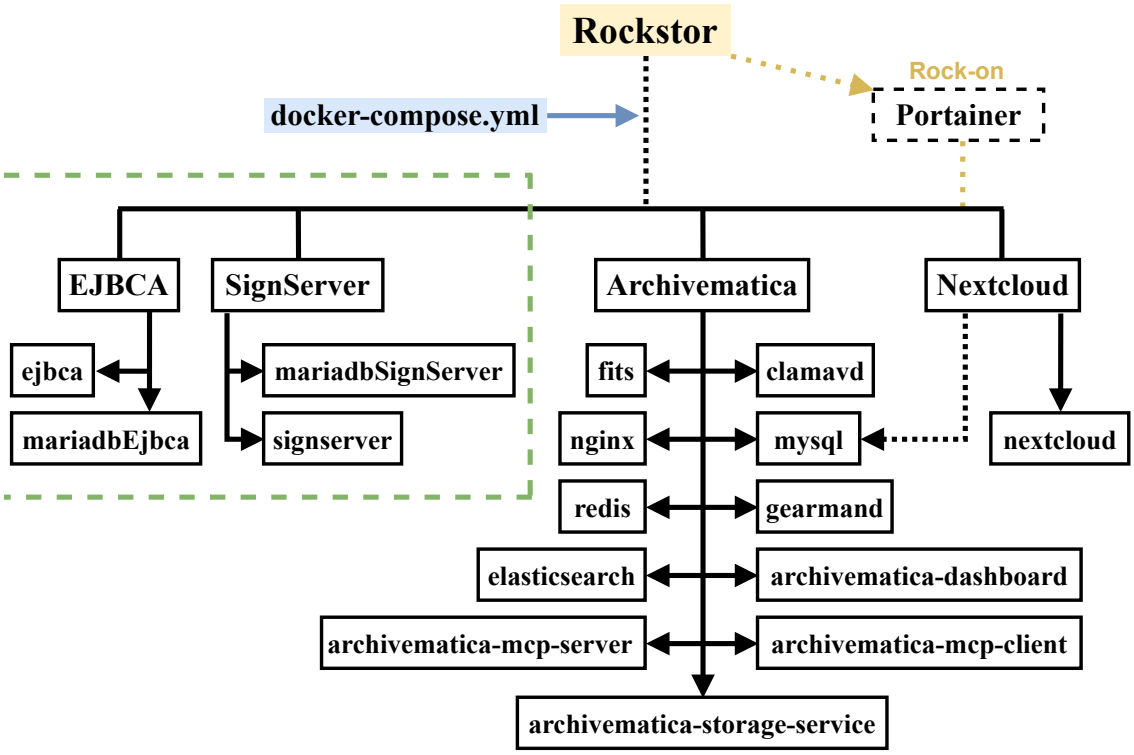
Zavádzanie kontajnerov na server

V prípade potreby podpisovania lokálne na tom istom serveri, je možné pridať do docker-compose súboru aj certifikačnú autoritu vo forme Ejbca (predpripravený obraz *"primekey/ejbca-ce"*) a podpisový server SignServer (predpripravený obraz *"primekey/signserver-ce"*). Pre obe služby je potrebné vytvoriť samostatné kontajnery s databázou (predpripravený obraz *"mariadb"*), kvôli bezpečnosti a samostatnosti jednotlivých serverov.

Výpis 9.2: Časť docker-compose súboru pre konfiguráciu kontajnera Nextcloud.

```
nextcloud:
    image: "nextcloud"
    ports:
      - "9080:80"
    environment:
      ADMIN_USER: "admin meno pre nextcloud"
      ADMIN_PASSWORD: "admin heslo pre nextcloud"
      MYSQL_HOST: "nazov kontajnera databazy"
      MYSQL_USER: "pouzivatel databazy"
      MYSQL_PASSWORD: "heslo do databazy"
      MYSQL_PORT: "port databazy"
    depends_on:
      - "nazov kontajnera databazy"
    links:
      - "nazov kontajnera databazy"
    volumes:
      - "nextcloud:/var/www/html"
      - "../src/archivematica-sampledata/:/home/archivematica
         /archivematica-sampledata/:rw"
      - "archivematica_pipeline_data:/var/archivematica/
         sharedDirectory/:rw"
      - "./mnt2/data-secure-luks2/:/data-secure-luks2/:rw"
      - "./mnt2/data-secure-luks1/:/data-secure-luks1/:rw"
```

Upozornenie: Pri vytváraní a dopĺňaní obsahu docker-compose súboru, treba dbať na presnú syntax používanej verzie, ako aj vyvarovaniu sa použitia zbytočných medzier, či tabulátora. Samotný docker engine síce pri použití kontroluje štruktúru obsahu, ale nie vždy dobre identifikuje miesto chyby (presný riadok), alebo príčinu, kedy môže poukazovať na chybu syntaxu (error code), ktorá tam nie je a problém je úplne v niečom inom. Dôležité je aj mapovanie portov aby nevznikali kolízie v smerovaní. Pri pridávaní nových umiestnení (volumes), je potrebné určiť všetky požadované prístupové práva daným službám podľa ich zamerania. Taktiež sa pri reálnom nasadení po fáze testovania odporúča všetky prihlasovacie údaje z docker-compose premiestniť do samostatného zabezpečeného oddielu (Docker secrets), odkiaľ budú jednotlivo načítané.



Obr. 9.5: Výsledná hierarchia kontajnerov.

9.6 Zavádzanie kontajnerov na server

Pred vytvorením kontajnerov treba na systém Rockstor doinštalovať niektoré nástroje. Ide o docker-compose, python-devel, git, a pokiaľ sa ráta s využitím GPG šifrovaných umiestnení na Archivematike tak aj `rng-tools`, kde následne kontajner Archivematica Storage Service musí mať zabezpečený prístup do `/dev/random`. Na vyhradené umiestnenie pre docker kontajnery je potrebné naklonovať repozitár aj s ďalšími submodulmi.

```
$ git clone https://github.com/artefactual-labs/am.git
$ git submodule update -init -recursive
```

Následne v zložke compose je potrebné upraviť súbor `docker-compose.yml` podľa návrhu do finálnej podoby, viď obr. 9.5, alebo ho rovno vymeniť za predpripravený. Príkazom `make create-volumes` sa zo súboru Makefile vykonajú inštrukcie, ktoré vytvoria potrebné umiestnenia pre budúce kontajnery Archivematiky. V tomto bode už je všetko pripravené na konečné vytváranie skupiny kontajnerov. Pomocou príkazu `docker-compose up -d -build` sa stiahnu a zostavia všetky potrebné obrazy, z ktorých sa následne vytvoria a spustia kontajnery podľa nastavení v docker-compose.yml ako jedna skupina (stack). Dĺžka tohto procesu závisí od výkonu dostupného hardvéru, ako aj od rýchlosti internetového pripojenia. Po úspešnom zosta-

vení skupiny, príkazom make bootstrap sa vykonajú zo súboru Makefile dodatočné nastavenia pre Archivematiku, po ktorých treba všetky jej kontajnery reštartovať.

```
$ make create-volumes
$ docker-compose up -d -build
$ make bootstrap
$ make restart-am-services
```

V základe sa spustí iba jeden kontajner archivematica-mcp-client pre vykonávanie úloh. Ak to výkon zariadenia dovoľuje, je možné jednoducho úpravou príkazu docker-compose vytvorenie viacero týchto služieb, ktoré zrýchlia archivačný proces systému.

```
$ docker-compose up -d -scale archivematica-mcp-client=3
```

Upozornenie: Príkaz make bootstrap sa niekedy nespustí na prvý krát. Tak isto sa nemusí hneď podariť vykonať hneď všetky akcie. V takom prípade treba príkaz aj niekoľkokrát zopakovať. Následne po reštartovaní kontajnerov sa nie vždy musia všetky spustiť. Príkladom je kontajner Archivematica dashboard, ktorý niekedy treba manuálne naštartovať aj trikrát za sebou kým sa korektne spustí. Pri týchto problémoch je výhoda správy cez Portainer, kde sa problémové kontajnery v stacku poľahky jednotlivo reštartujú alebo nanovo spustia.

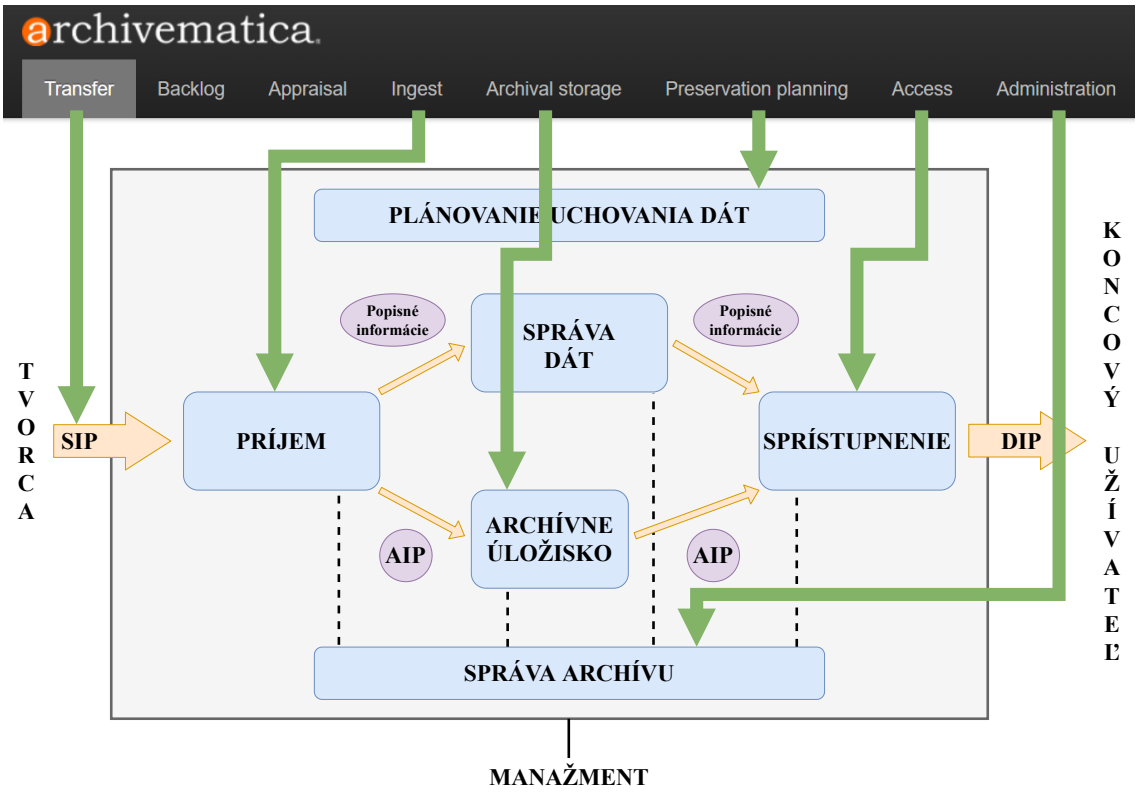
9.7 Nastavenie a funkcionalita Archivematiky

Najdôležitejšia časť pre používateľa je webové rozhranie (dashboard). Predvolené prihlasovacie údaje pre testovacie účely sú: používateľské meno (test), heslo (test). Po prihlásení sa zobrazí menu, ktoré predstavuje jednotlivé funkčné entity modelu OAIS, viď obr. 9.6.

V archivačnom systéme Archivematica je proces prenosu a prípravy digitálnych objektov označovaný ako Transfer. Na tejto karte používatelia vyberú materiál, ktorý sa má archivovať, pomenujú ho a zahája proces prenosu a vytvárania balíkov SIP.

Na výber sú nasledujúce možnosti:

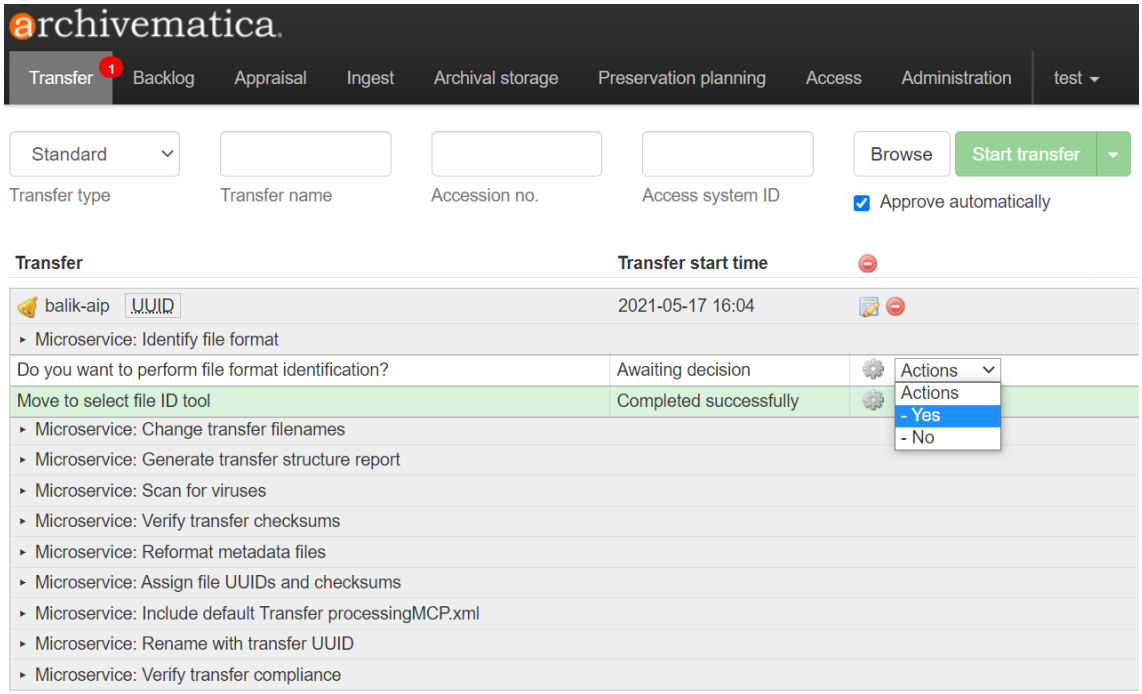
- **Transfer type** – druh prenášaného materiálu.
  - Standard – štandardné prevody sú predvolené. Všetky materiály je možné prenášať pomocou tohto typu prenosu. Nevykonávajú sa žiadne špeciálne úlohy spracovania.
  - Zipped directory – materiály sú nahrávané zabalene ako `.zip, .tgz, .tar.gz`. Pri spustení prenosu takéhoto adresára, sa najskôr rozbalí (dekomprimuje) a následne spracuje ako v type Standard.



Obr. 9.6: Prepojenie archivematiky na OAIS model.

  - Unzipped bags – archivačný systém vie rozoznať typ a štruktúru predpripraveného adresára podľa štandardu balíkov BagIt. Dokáže takýto balík overiť kontrolou súboru manifest a vypočítaných kontrolných súčtov.
  - Zipped bags – balíky podľa štandardu BagIt, sú nahrávané zabalene ako `.zip, .tgz, .tar.gz`. Pri spustení prenosu takéhoto balíka, sa najskôr rozbalí (dekomprimuje) a následne spracuje ako v type Unzipped bags.
  - Disk image – predpokladá prenos obrazu disku, kde poskytuje možnosť pridania ďalších špecifických metadát.
  - DSpace – používa sa pri exporte materiálov z repozitára DSpace.
  - Dataverse – používa sa pri exporte materiálov z repozitára Dataverse.
- **Transfer name** – názov prenosu, ktorý sa stane názvom výsledného AIP balíka. Vyplnenie tohto poľa je povinné.
- **Accession no.** – zadanie prístupového čísla bude mať za následok jeho pridanie do súboru metadát METS balíka AIP. Nepoužíva sa na identifikáciu alebo vyhľadávanie. Toto pole je voliteľné.
- **Access system ID** – prístupové ID sa zadáva pre možnosť automatizácie procesu nahrávania balíkov DIP do špecializovaných aplikácií pre prístup k obsahu. Toto pole je voliteľné.

- **Browse** – tlačidlo na otvorenie prehliadača súborov, ku ktorým má archivačný systém prístup, a sú označené ako zdroj pre archiváciu (prenos). Výberom adresára a potvrdením tlačidlom Add sa dané materiály pridajú do nastavení prenosu.
- **Start transfer** – po nastavení všetkých potrebných náležitostí sa po stlačení tlačidla začne proces prenosu a prípravy balíka SIP pomocou mikroslužieb.
- **Approve automatically** – ak pole nie je zaškrtnuté, po spustení prenosu sa proces zastaví pri prvej mikroslužbe, a bude vyžadovať manuálne potvrdenie pre pokračovanie.
- **Processing configuration** – pri tlačidle Start transfer je rolovacie menu (šípka nadol), v ktorom sa zobrazuje možnosť výberu konfiguračného profilu pre spracovanie obsahu. Ak používateľ nezvolí žiaden, použije sa predvolený (default).



Obr. 9.7: Manuálne rozhodovanie pri procese archivácie.

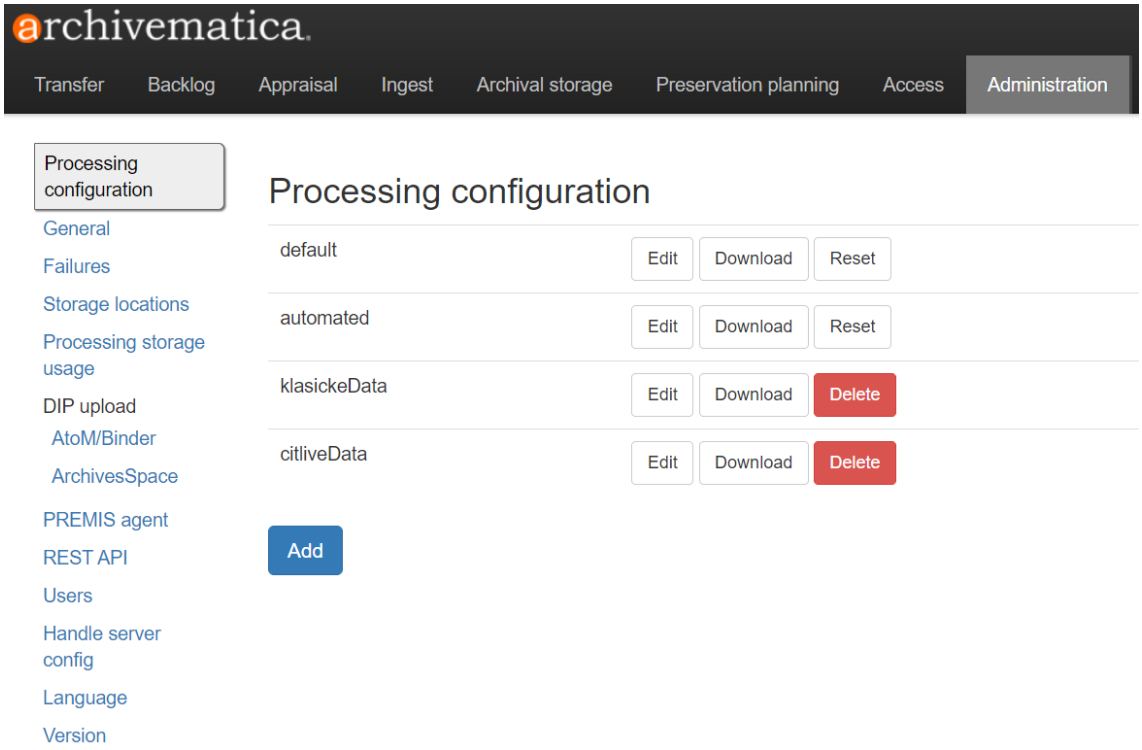
9.7.1 Automatizácia procesu archivácie

Výberom konfiguračného profilu pri štarte procesu prenosu (Transfer) sa vykonajú určité úlohy, ktoré sú v systéme Archivematica implementované formou mikroslužieb. Tie sa delia na dve základné sekcie:

1. **Mikroslužby pre Transfer** – vykonávajú prenos, potvrdenie a kontrolu obsahu pre archiváciu. Následne vytvoria balík SIP.

2. **Mikroslužby pre Ingest (príjem)** – vykonávajú kontrolu SIP balíka, normalizáciu obsahu, pridanie potrebných metadát a následne vytvorenie a uchovanie AIP a DIP balíkov.

Všetky úlohy (mikroslužby) sú rozdelené a popísané pri každom jednom procese archivácie zvlášť v karte Transfer a Ingest. Je možná ich jednotlivá kontrola, ako aj používateľsky zasiahnuť do ich procesu, pretože viaceré z nich potrebujú vybrať nejakú možnosť pre určenie ako majú narábať s obsahom v daný moment (body rozhodovania), viď obr. 9.7. Takýto proces je zdĺhavý a pri väčšom počte archivovaných dát neefektívny. Archivematica preto umožňuje čiastočne alebo celkovo zautomatizovať archiváciu.



Obr. 9.8: Vytvorené konfigurácie pre automatický proces archivácie.

Nastavenie automatizácie je možné v karte Administration pod Processing configuration. Po inštalácií sú dostupné dve základné konfigurácie. Default sa používa pri klasickom stlačení tlačidla Start transfer, a automated je automatizovaná voliteľná možnosť. Obe sa dajú editovať. Správca môže vytvoriť viacero týchto procesných konfigurácií, pre rozdielne typy vstupných dát, viď obr. 9.8. Používateľ si potom len pri počiatočnom nastavení v karte Transfer pri tlačidle Start transfer vyberie potrebnú možnosť z rolovacieho menu.

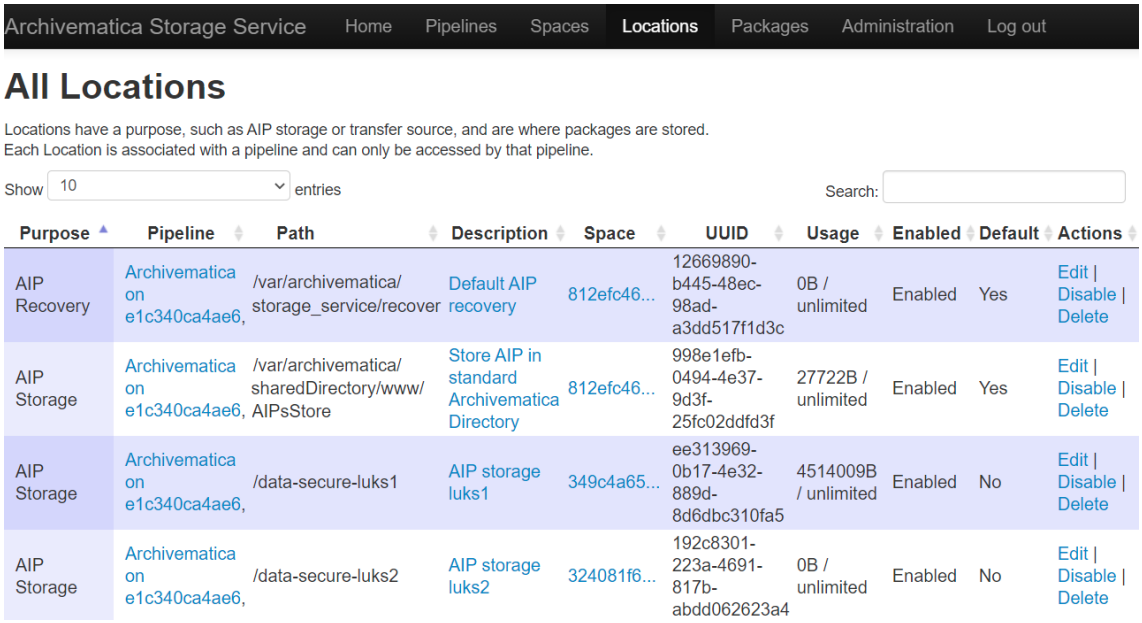
V tomto riešení je základom vytvorenie aspoň dvoch dodatočných konfigurácií, ktoré budú taktiež plne automatizovať proces archivácie a to pre klasické dáta

a citlivé dáta. Hlavným rozdielom je lokácia uchovania balíkov AIP, a taktiež žiadna kompresia citlivých dát. Pri bežných dátach sa využíva kompresný algoritmus LMZA pomocou 7z, a úroveň je nastavená normálna (5 – normal compression) pre najväčšiu rovnováhu medzi rýchlosťou a veľkosťou.

V karte Administration pod General, je dôležité nastavenie algoritmu kontrolných súčtov. Ten je potreba zvoliť vždy čo najbezpečnejší (v tomto prípade je to SHA-512), pretože hlavne pri dlhodobej archivácií sa nedá predpokladať kedy budú jednotlivé algoritmy považované za prelomiteľné.

9.7.2 Archivematica Storage Service

Vybrané umiestnenia ku ktorým má Archivematica prístup sú definované pri vytváraní a nasadzovaní kontajnerov v súbore docker-compose. Následne ich treba manuálne pridať priamo do Archivematica Storage Service, ktorá je dostupná zo samostatného webového rozhrania, viď obr. 9.9.



Obr. 9.9: Zobrazenie lokácie AIP balíkov v Storage Service.

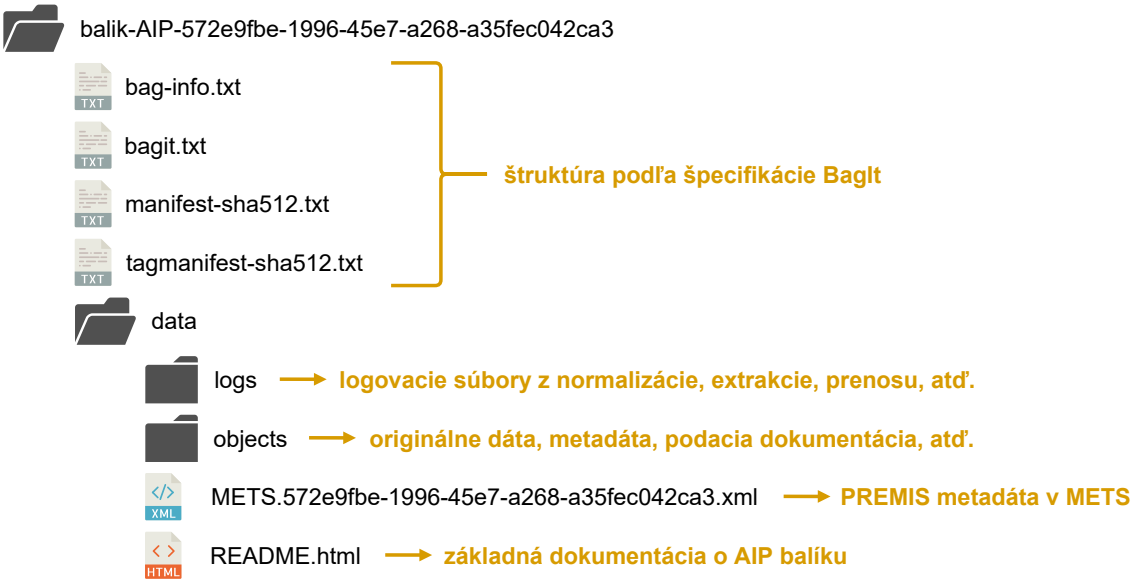
V karte Spaces je v základe definovaný lokálny priestor, kde sa ukladajú informácie. Tu je potrebné vytvoriť ďalšie dve umiestnenia s rovnakým prístupovým protokolom (Local Filesystem), ale absolútna cesta bude rozdielna pre bežné a citlivé dáta (data-secure-luks1, data-secure-luks2). Následne je potreba pri každom z nich vytvoriť novú lokáciu (Create location here) za účelom (purpose) uloženia AIP (AIP storage). Zvolením relatívnej cesty v danom priestore sa určí presný adresár, do ktorého sa balíky budú ukladať. Úspešné vytvorenie lokácie si je možné skontrolovať

v karte Locations. Vytvorené nové umiestnenia sa pridali ako ďalšia možnosť výberu úložiska AIP pri procese archivácie v samotnej Archivematike.

Storage Service dokáže vytvoriť umiestnenie, ktoré bude šifrované pomocou GPG. Potrebné kľúče je možné vytvoriť alebo naimportovať v karte Administration v sekcii Encryption keys. Šifrovanie balíkov AIP ale predstavuje dodatočne značné zvýšenie potreby výpočtového výkonu a preto sa neodporúča používať na pomalších zariadeniach.

9.7.3 Štruktúra AIP

Výsledný archivačný balík v systéme Archivematica pozostáva zo základu podľa špecifikácie BagIt, ku ktorému sú pridané logy o vykonaných akciách pri archivácií, viď obr. 9.10. Ďalšími položkami sú metadáta, kde je podporovaný štandard PREMIS (verzie 3), ktorý je vnorený do METS. Výsledný súbor tak zhromažďuje rozsiahle technické metadáta o objektoch a zaznamenáva vykonané akcie (známe ako udalosti).



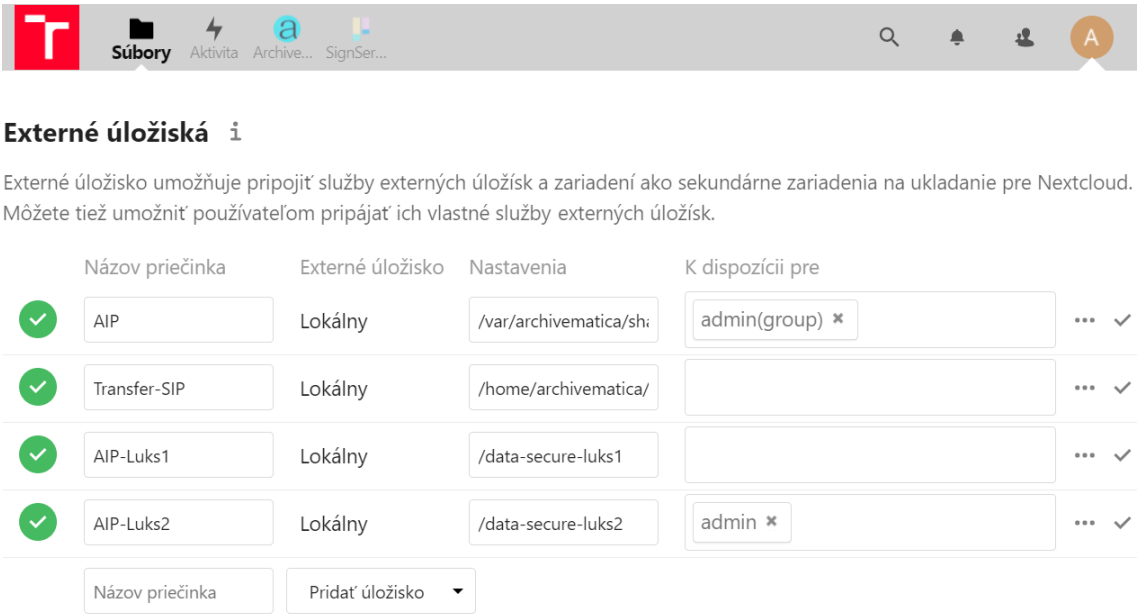
Obr. 9.10: Štruktúra AIP balíka v systéme Archivematica.

Práve tento súbor METS vo formáte XML je možné stiahnuť v sekcií Archival storage po vybratí potrebného AIP kliknutím na tlačidlo View. Kvôli jeho komplexným informáciám o balíku (o dátach, procese archivácie, atď.) sa odporúča jeho úschova na externé úložisko, kvôli kontrole dát vyextrahovaného AIP v budúcnosti. Taktiež je možné tento súbor obstarať digitálnym podpisom pre zaručenie autenticity.

9.8 Nastavenie a funkcionalita Nextcloudu

Nextcloud predstavuje pre používateľa hlavný nástroj, na ktorom bude uchovávať a manažovať svoje dáta. Poskytuje širokú škálu možností nastavenia aj pomocou aplikácií, ktoré sú buď vstavané, alebo ich je jednoduché stiahnuť z ponuky. Základnými potrebnými rozšíreniami pre toto riešenie sú aplikácie:

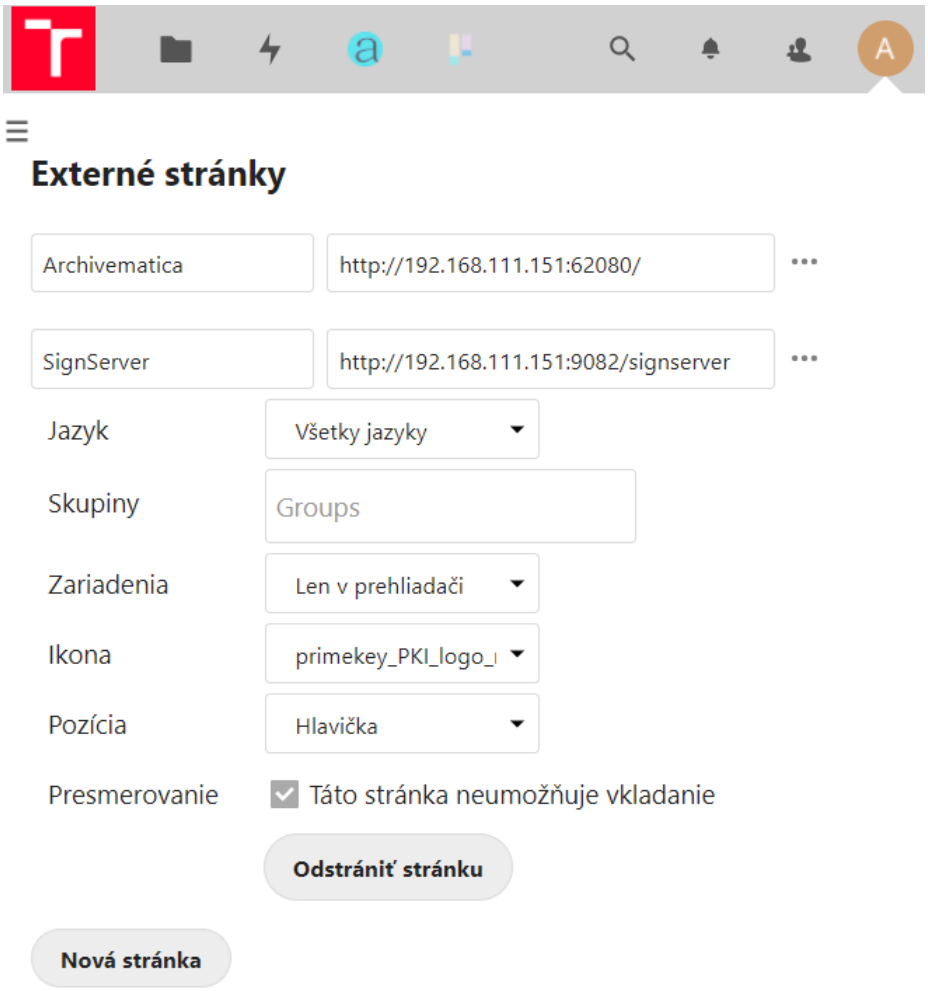
- **External storage support** – táto aplikácia je predinštalovaná, treba ju len zapnúť v nastaveniach. Umožňuje správcovi nakonfigurovať prepojenia k externým ukladacím priestorom ako sú servery FTP, SFTP, atď., ale aj lokálne úložiská, ktoré sú mimo samotného kontajnera. Možnosť nastavenia je v sekcií Administrácia v karte Externé úložiská, viď obr. 9.11. Tu je potrebné pridať lokálne úložiská, ktoré Archivematica Storage Service používa ako zdroj pre archiváciu, a tak isto úložiská kde ukladá vytvorené AIP balíky. Tie sa následné zobrazia v prehliadači súborov Nextcloud ako samostatné adresáre. Dôležité je aj nastavenie prístupu z daným umiestneniam, kde pre citlivé dáta sa odporúča povoliť prístup len overeným používateľom jednotlivo.



Obr. 9.11: Nastavenie externých úložísk v Nextcloude.

- **External sites** – táto aplikácia nie je predinštalovaná a je potrebné ju stiahnuť a povoliť. Je dostupná v nastavení aplikácií v sekcií prispôsobenie. Táto aplikácia umožňuje správcovi pridávať ďalšie odkazy (externé webové stránky) do horného menu Nextcloudu. Po kliknutí na prepojené stránky sa zobrazia ako súčasť otvoreného okna. Pre pridanie je potrebné je vyplniť názov, ktorý sa bude zobrazovať v menu a webový odkaz (prípadne IP adresu), viď obr. 9.12.

Je možné obmedziť, ktorým používateľom sa budú tieto odakzy zobrazovať a na akom type zariadení. V tomto riešení je potrebné pripojiť Archivematiku a SignServer. Niektoré stránky majú z bezpečnostných dôvodov zablokovanú funkciu iframe. Tak je to aj v prípade SignServeru, kde je potrebné zaškrtnúť možnosť *Táto stránka neumožňuje vkladanie*. Stránka sa teda nebude zobrazovať priamo v okne Nextcloud, ale otvorí sa ako samostatná karta prehliadača.



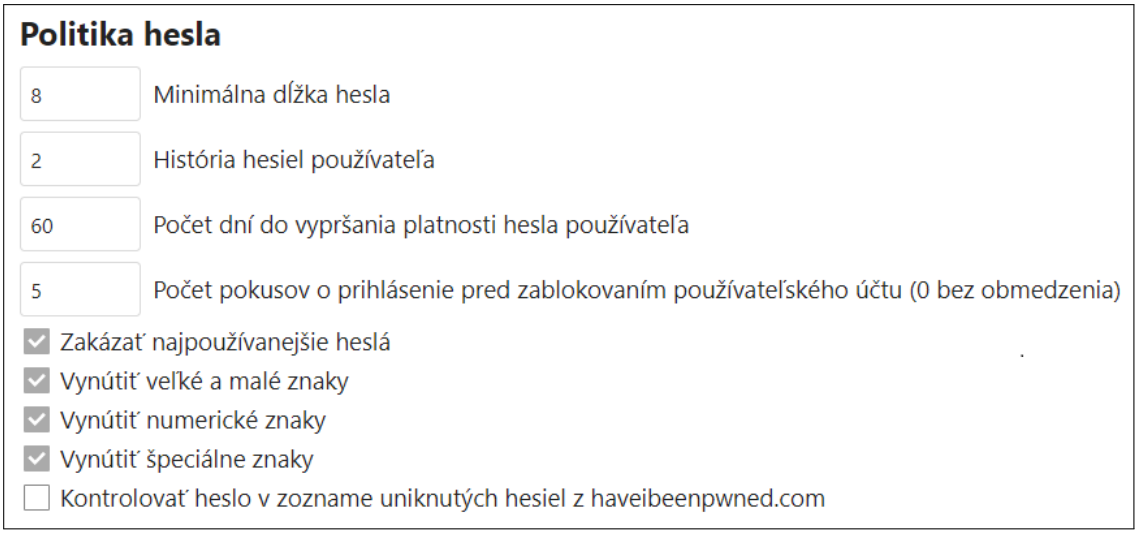
Obr. 9.12: Nastavenie externých stránok v Nextcloude.

9.8.1 Bezpečnosť Nextcloudu

Nextcloud ponúka širokú škálu bezpečnostných nastavení vstavanými ako aj externými aplikáciami. Najdôležitejšie sú v oddiele Administrácia v sekcii Zabezpečenie.

Šifrovanie obsahu na serveri Nextcloud je možné, ale v základe je vypnuté. Pre použitie v tomto riešení sa taktiež odporúča nechať neaktívne, pretože táto funkcia nešifruje externe pridané úložiská, kde sa uchovávajú balíky AIP. Taktiež podstatne zvyšuje potrebu výpočtového výkonu a zväčšuje veľkosť uchovávaných dát až o 35 %.

Ďalšou bezpečnostnou funkciou je politika hesiel, kde administrátor môže obmedziť minimálnu dĺžku hesla ako aj vynútenie použitie špeciálnych znakov, viď obr. 9.13. Čím je v hesle viacero odlišných znakov (čísla, veľké a malé znaky, atď.), tým môže byť heslo kratšie a stále sa považuje za bezpečné.



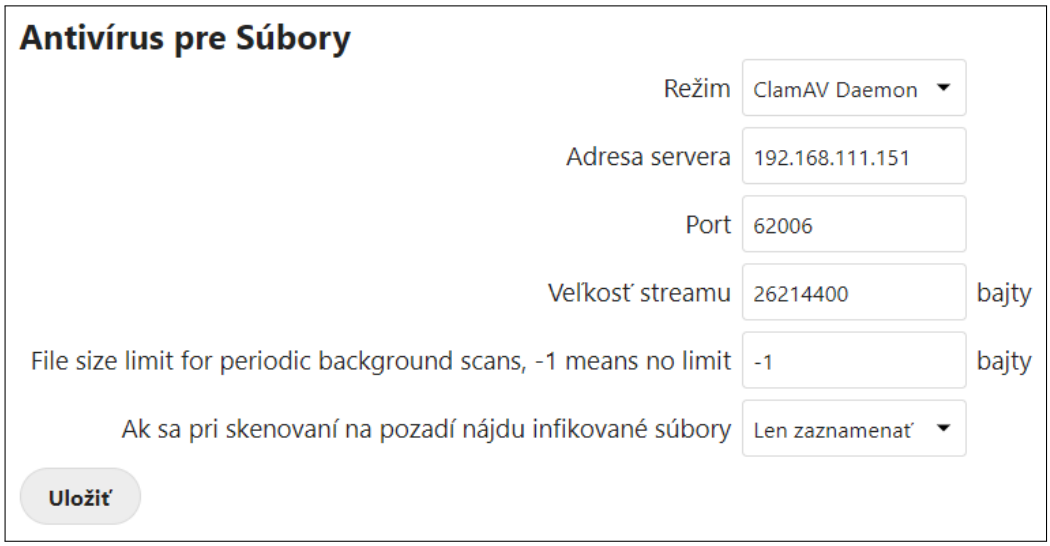
Obr. 9.13: Nastavenie politík hesiel v Nextcloude.

Štandardne Nextcloud povoľuje prihlasovanie z ľubovoľnej IP adresy. Administrátor môže prihlasovanie obmedziť len na určité rozsahy IP adries. Taktiež má automaticky zapnutú ochranu pred útokmi hrubou silou, ale pre testovacie účely je ale možné povoliť rozsah adries, ktoré nebudú blokované.

Poslednou hlavnou základnou bezpečnostnou funkciou je antivírus pre súbory. Využíva funkcionalitu ClamAV alebo Kaspersky (len ako daemon). V pokročilých nastaveniach je možné upraviť predvytvorené pravidlá skenera ako aj vytvoriť vlastné. Dokáže bežať v režimoch:

- **Executable** – ClamAV beží na rovnakom serveri ako Nextcloud a pri každom načítaní súboru sa spustí a zastaví príkaz clamscan. Je nutná dodatočná inštalácia ClamAV, kedy by mal Nextcloud jeho umiestnenie následne sám detekovať.
- **Daemon** (Socket) – ClamAV beží na rovnakom serveri, kde beží na pozadí ako daemon a využívajú ho aj iné aplikácie. Nextcloud by mal detekovať clamd socket automaticky. Túto možnosť v kontajnerovom prostredí nie je dobré používať, pretože v základe sa kontajnery správajú ako samostatné jednotky a detekcia teda nebude fungovať.
- **Daemon** – ClamAV alebo Kaspersky beží na inom serveri. Pre využitie je potrebné zadať IP adresu servera a potrebný port ktorý má pridelená táto služba.

V tomto riešení sa využíva režim Daemon. Služba antivírusu ClamAV už beží ako samostatný kontajner pre potreby archivačného systému. Preto je jednoduché nastavenie IP adresy a portu príslušného kontajnera, viď obr. 9.14. Ušetrí sa tak výpočtový výkon. Ak sa nájdu infikované súbory, tak sa daná akcia zaznamená a upozorní sa správca.



Obr. 9.14: Konfigurácia antivírusu v Nextcloude.

9.8.2 Aktivita, logovanie a monitoring

V sekcií Aktivita má správca možnosť kontroly aktivity všetkých používateľov. Všetky zmeny v súboroch (nahratie, vymazanie, zdieľanie, vytvorenie, atď.) sú zaznamenané aj s časovými údajmi. Dostupné sú tu aj záznamy antivírusu. Podrobné serverové logy sú dostupné v nastaveniach v sekcií Logovanie. Každý záznam je zviazaný s časom kedy daná akcia nastala. Logy sa dajú prehľadne filtrovať ako aj stiahnuť v jednom súbore (nextcloud.log). V sekcií Systém sa nachádza monitoring systému. Správca tu môže sledovať vyťaženie systémových zdrojov ako aj kapacitu dostupných úložísk či počet aktívnych používateľov.

9.8.3 Ďalšie doplňujúce aplikácie

Nextcloud ponúka rôzne doplňujúce aplikácie. Vybrané z nich sú doporučené pre použitie aj v tomto riešení.

Extract – dokáže extrahovať súbory zabalené vo formáte zip, rar, tar, gzip, 7z, deb a bzip2. Potrebné je doinštalovanie PHP rozšírenia rar, a Linux balík p7zip. Použitie je vhodné na extrakciu obsahu uložených AIP balíkov priamo v Nextcloude.

File access control – umožňuje správcom chrániť údaje pred neoprávneným prístupom alebo úpravami. Správca môže vytvoriť a spravovať rôzne skupiny pravidiel. Použitie je dobré pre dôkladnú správu archivačných balíkov ako aj dát pre archiváciu. Príkladom môže byť zamedzenie prístupu k AIP úložisku určitým používateľom, alebo obmedzenie veľkosti súborov pridávaných do priečinka pre archiváciu.

Checksum – aplikácia dokáže vypočítať kontrolný súčet ľubovoľného súboru algoritmom MD5, SHA1, SHA256, SHA384, SHA512 alebo CRC32. Stačí otvoriť zobrazenie podrobností súboru (bočný panel), kde sa objaví nová karta s názvom Checksum. Následne už len stačí vybrať algoritmus a kontrolný súčet sa ihneď vygeneruje. Táto funkcia je výborná pre manuálnu kontrolu dát po extrakcií AIP balíka, kde sa týmto spôsobom odhalia nezrovnalosti s vypočítanými kontrolnými súčtami uloženými v metadátach. Taktiež je možné vytvoriť kontrolné súčty k súborom určeným pre archiváciu, ktoré následne spracuje archivačný systém.

Metadata – umožňuje zobrazenie dostupných metadát obrázkov, videí a audia. Podporuje najpoužívanejšie formáty ako jpeg, tiff, mp4, wav, flac a iné. Ponuka sa rozširuje každou aktualizáciou. Zobrazené metadáta sa dajú použiť ako doplnkové údaje pri procese archivácie pre identifikáciu obsahu.

LibreSign – táto aplikácia dokáže podpisovať PDF súbory. Ak je potrebné opatriť nejaké dokumenty pred archiváciou digitálnym podpisom, touto cestou je to možné priamo v Nextcloude. Potrebné je ale dosť upraviť samotný obraz kontajneru, kedy je potrebné doplniť určité rozšírenia do Dockerfile (java, JSignPDF). Keďže využíva sadu nástrojov cfssl, je nutné vytvoriť ďalší kontajner pre službu cfssl (doplnenie do docker-compose.yml). Následne v menu Nextcloudu je potrebné vyplnenie údajov certifikačnej autority pre root certifikát. LibreSign vyžaduje pre svoj chod množstvo zmien v systéme, preto sa odporúča jej použitie len skúseným správcom.

9.9 Popisový server SignServer

Je využívaný ako komunitná edícia verzie 5.2.0 (SignServer-CE 5.2.0). Primárne použitie v tomto riešení má pre podpisovanie METS súborov štandardom XAdES-T, prípadne PDF dokumentov.

Pre vykonávanie používateľských požiadaviek SignServer potrebuje vytvoriť pre každú funkciu takzvanú vykonávaciu entitu (Worker), viď obr. 9.15. Súčasťou systému sú aj predpripravené vzory týchto entít (zložka `doc/sample-configs/`), ktoré sa dajú upraviť podľa vlastnej potreby, alebo použiť priamo pre testovacie účely. Potrebné nastavenie je možné vykonať buď prostredníctvom príkazového riadku alebo cez prehľadné webové rozhranie.

Upozornenie: Administrátorské webové rozhranie je implementované do komunitnej edície práve od verzie 5.2.0, preto má veľa nedostatkov. Hlavný je, že ani podľa nastavenia z dokumentácie sa nedá často dostať do tohto webového rozhrania ani s potrebnými certifikátmi. Používateľ je tak odkázaný na konfiguráciu cez príkazový riadok, čo podstatne spomaľuje a znemožňuje prehľadné nastavenie.

Ako prvé treba nastaviť worker pre Crypto Token, ktorý poskytuje prístup ku kľúčom a kryptografickým operáciám. Ten v základe využíva softvérové úložisko kľúčov PKCS#12 (.p12/.pfx), ale zmenou v konfiguračnom súbore keystore-crypto.properties dokáže používať aj Java JKS (.jks) v lokálnom súborovom systéme. Ak je dostupný HSM modul, alebo jeho softvérová náhrada softHSM, stačí upraviť súbor pkcs11-crypto.properties a vytvoriť worker priamo s ním. Vytvorenie CryptoWorkera založenom na PKCS#12 (CryptoTokenP12). Po každej konfigurácií sa musí SignServer znovu načítať pre uplatnenie zmien. Cesta ku keystore súboru a základné heslo sa musia nastaviť ešte pred vytvorením v konfiguračnom súbore.

```
$ bin/signserver setproperties doc/sample-configs/
keystore-crypto.properties
$ bin/signserver reload 1
```

Podobným spôsobom je potrebné vytvoriť worker pre služby časovej pečiatky (TimeStampSigner). Predtým je ale ešte nutné nastaviť základný kľúč do vytvoreného úložiska kľúčov (p12).

```
$ bin/signserver setproperty 1 KEYSTOREPATH $SIGNSERVER_HOME/cesta
/k/suboru/keystore.p12
$ bin/signserver setproperty 1 KEYSTOREPASSWORD heslo_keystore
$ bin/signserver setproperty 1 DEFAULTKEY "heslo_ts"
$ bin/signserver reload 2
$ bin/signserver setproperties doc/sample-configs/timestamp.properties
$ bin/signserver reload 3
```

Ďalší worker je potrebný vytvoriť pre modul podpisovania PDF (PDFSigner).

```
$ bin/signserver setproperty 1 KEYSTOREPATH $SIGNSERVER_HOME/cesta
/k/suboru/keystore.p12
$ bin/signserver setproperty 1 KEYSTOREPASSWORD heslo_keystore
$ bin/signserver setproperty 1 DEFAULTKEY "heslo_signerPDF"
$ bin/signserver reload 4
$ bin/signserver setproperties doc/sample-configs/pdfsigner.properties
$ bin/signserver reload 5
```

Posledný worker je pre modul podpisovania XML súborov štandardom XAdES (XAdESSigner). Ten je ale v základe nastavený na štandard XAdES-BES, ale toto rie-

šenie vyžaduje použitie aj časovej pečiatky, štandard XAdES-T. V konfiguračnom súbore je potrebné nastaviť:

```
WORKERGENID1.XADESFORM=T
WORKERGENID1.TSA_WORKER=TimeStampSigner
WORKERGENID1.TSA_URL=
http://ip-adresa:8080/signserver/tsa?workerName=TimeStampSigner
```

Následne sa už len vytvorí základné heslo a worker.

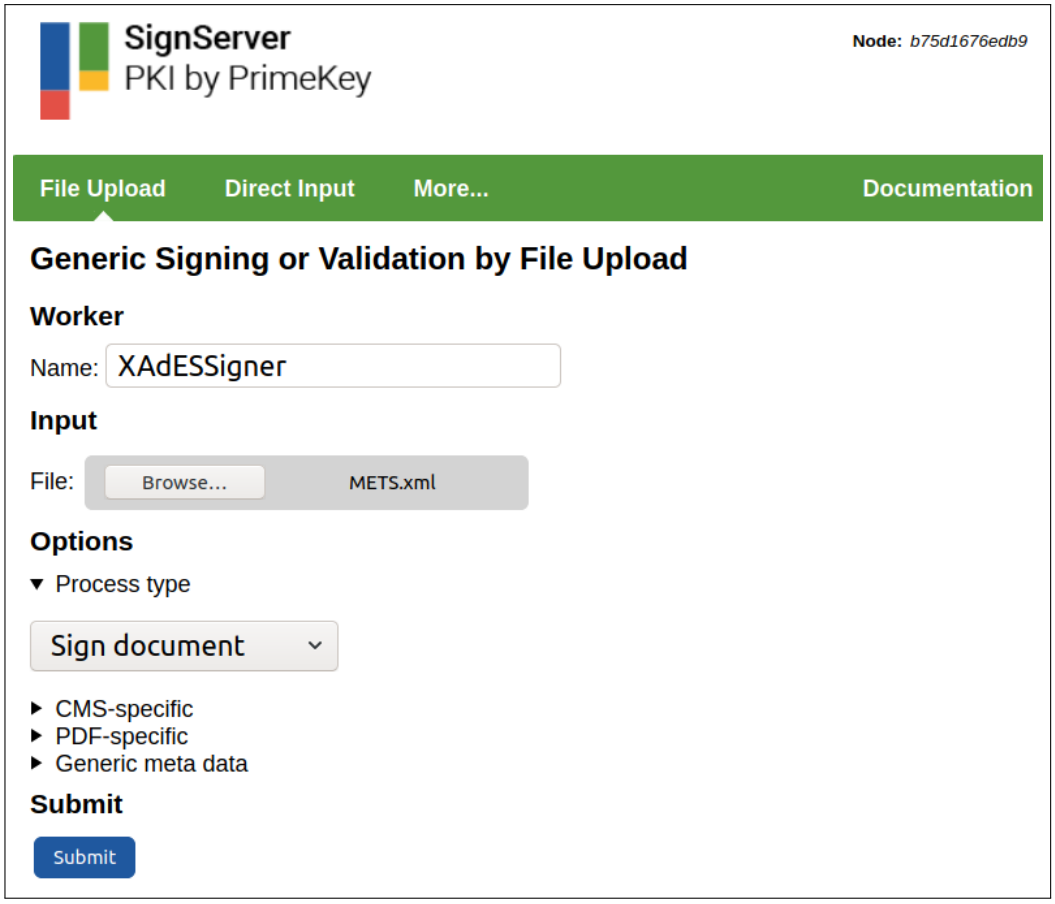
```
$ bin/signserver setproperty 1 KEYSTOREPATH $SIGNSERVER_HOME/cesta
/k/suboru/keystore.p12
$ bin/signserver setproperty 1 KEYSTOREPASSWORD heslo_keystore
$ bin/signserver setproperty 1 DEFAULTKEY "heslo_signerXML"
$ bin/signserver reload 6
$ bin/signserver setproperties doc/sample-configs
/xadessigner.properties
$ bin/signserver reload 7
```

```
[root@b75d1676edb9 signserver]# bin/signserver getstatus brief all
2021-05-19 INFO [naming] WildFly Naming version 1.0.9.Final
2021-05-19 INFO [security] ELY00001: WildFly Elytron version 1.6.0.Final
2021-05-19 INFO [threads] JBoss Threads version 2.3.2.Final

Current version of server is : SignServer CE 5.2.0.Final

Status of CryptoWorker with ID 1 (CryptoTokenP12) is:
   Worker status : Active
   Token status  : Active

Status of Signer with ID 2 (TimeStampSigner) is:
   Worker status : Active
   Token status  : Active
   Signings      : 0 (counter disabled)

Status of Signer with ID 3 (PDFSigner) is:
   Worker status : Active
   Token status  : Active
   Signings      : 0 (counter disabled)

Status of Signer with ID 4 (XAdESSigner) is:
   Worker status : Active
   Token status  : Active
   Signings      : 0 (counter disabled)
```

Obr. 9.15: Výpis aktívnych workerov.

Použitie týchto workerov je možné cez príkazový riadok alebo v používateľskom webovom rozhraní, kde sa zadáva presné meno workera (napríklad XAdESSigner), viď obr. 9.16. Vstupom je potom súbor, ktorý je potrebné podpísať. Po stlačení tlačidla Submit, worker vykoná požiadavku a podpíše daný súbor.

SignServer je výborný nástroj na podpisovanie dát, ale v komunitnej verzií má značne obmedzenú funkcionalitu. Taktiež pre dôkladné nastavenie vyžaduje dosť času a skúseného správcu, pretože dostupná dokumentácia nepostačuje a komunitná podpora je len v obmedzenej podobe.



Obr. 9.16: Webové používateľské rozhranie SignServera.

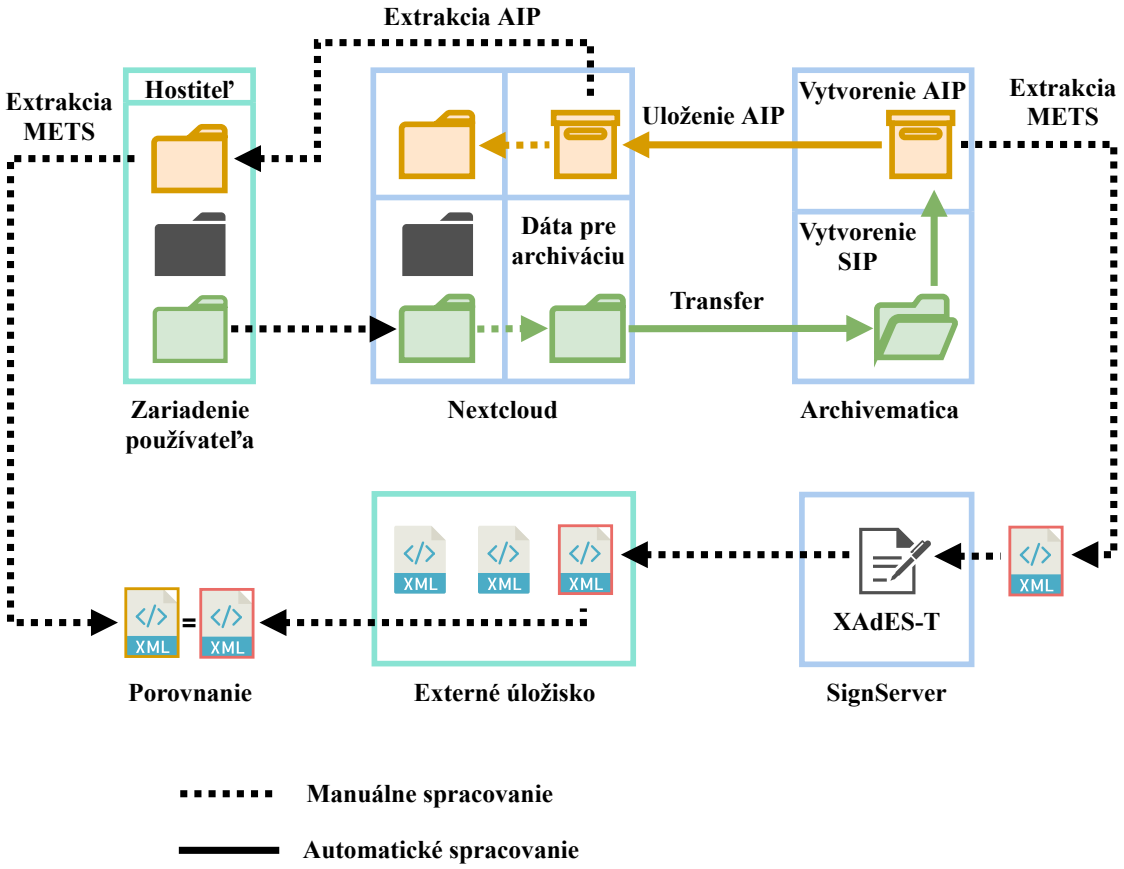
9.9.1 Ejbca

Je najviac voliteľným prvkom v tomto riešení. Zastáva miesto certifikačnej autority (CA) v systéme, ktorá dokáže poskytovať potrebné certifikáty podpisovému serveru. Príkladom môže byť vygenerovanie TLS certifikátu, alebo pre potreby RenewalWorkera, ktorý sa používa na generovanie nového páru kľúčov a obnovenie certifikátu workerov z EJBCA. SignServer vygeneruje CSR (Certificate Signing Request) a odošle ho CA. Tá vydá potrebný certifikát (vo formáte pem), ktorý sa následne naimportuje na podpisový server. Celý tento proces dokáže EJBCA automatizovať, ale iba v Enterprise verzii. Ovládanie a nastavenie celej CA je prehľadné prostredníctvom webového rozhrania.

Celkovo by sa certifikačná autorita nemala nachádzať na jednom serveri z bezpečnostných dôvodov, preto sa odporúča použitie externej overenej certifikačnej autority. Taktiež pri dlhodobej archivácií je dobré aby sa dali podpísané údaje overiť aj s odstupom času. Preto sú lokálne certifikačné autority nie moc vhodné, a považujú sa za posledné možné riešenie.

9.10 Funkcionalita z pohľadu používateľa

Z pohľadu bežného používateľa je funkcionalita a možnosti použitia výsledného riešenia nasledovná, viď obr. 9.17. Používateľ má na svojom zariadení dáta, ktoré si uloží na privátny cloud (Nextcloud). Neskôr sa rozhodne, že časť dát je potrebné archivovať. Presunie ich tak do oddielu dát pre archiváciu, do ktorého má prístup aj archivačný systém. Následne buď poverí zodpovednú osobu, alebo sa sám prihlási do archivačného systému (podľa nastavených politík a oprávnení). Taktiež, ak ide o citlivé dáta, odomkne úložisko v systéme Rockstor (manuálne odomknutie šifrovaných diskov pomocou LUKS 2).



Obr. 9.17: Schéma funkcionality z pohľadu používateľa.

V archivačnom systéme vyberie priečinok, ktorý chce archivovať a zvolí typ automatického spracovania podľa typu dát. Systém automaticky vykoná všetky potrebné procesy (vytvorenie SIP, AIP, atď.) a uloží výsledný archivačný balík AIP na prednastavené úložisko. Následne môže používateľ extrahovať metadáta METS a podpísať ich na podpisovom serveri (SignServer) štandardom XAdES-T. Takto podpísaný súbor metadát uloží na externé úložisko alebo iné ľubovoľné médium pre potreby kontroly integrity a autenticity archivovaných dát v budúcnosti. Po určitej dobe bude používateľ potrebovať časť z dát, ktoré predtým archivoval. Prihlási sa do privátneho cloudu, ktorý má prístup k umiestneniam AIP. Vyextrahuje dáta z potrebného balíka buď priamo na privátnom cloude, alebo na svojom zariadení. Aby si overil, že s obsahom nikto nedovolene nemanipuloval po dobu archivácie, porovná obsah súbor METS z extrahovaného balíka s jeho podpísaným ekvivalentom uloženým na externom úložisku.

9.11 Hardvérové nároky

Hardvérové nároky realizovaného riešenia závisia od počtu používateľov ako aj veľkosti dát s ktorými sa pracuje. V nasledujúcej tabuľke 9.2 sú zobrazené minimálne požiadavky na procesor (CPU) ako aj pamäť (RAM), jednotlivých použitých komponentov podľa dokumentácií. Tieto údaje znázorňujú potrebné hardvérové nároky pre základnú funkcionalitu. Napríklad dokumentácia Nextcloud udáva, že pri prostredí do 150 užívateľov sa odporúča až 16 GB pamäte. Takisto treba rátať s vyššími nárokmi pri rozšírení funkcionality pomocou podporných aplikácií či systémov.

Tab. 9.2: Hardvérové nároky komponentov, [58], [59], [67], [78].

	CPU	**RAM**	
	minimálne	**minimálne**	**doporučené**
Rockstor	2 jadrá	2 GB	4 GB
Archivematica	2 jadrá (1,0 + GHz)	2 + GB	4 + GB
Nextcloud	2 jadrá (1,0 + GHz)	128 MB	512 MB
SignServer	2 jadrá	1 GB	1,5 GB
EJBCA	2 jadrá	1 GB	1,5 GB

Návrh predpokladá nasadenie na jeden fyzický server. V tejto práci boli podmienky pre reálne nasadenie simulované prostredníctvom virtuálneho stroja, na ktorom bežal systém Rockstor. Pridelené hardvérové zdroje boli dve fyzické jadrá procesora `Intel Core i5-4670K` (takt jadier 4,5 GHz), a 8 GB pamäte RAM. Po

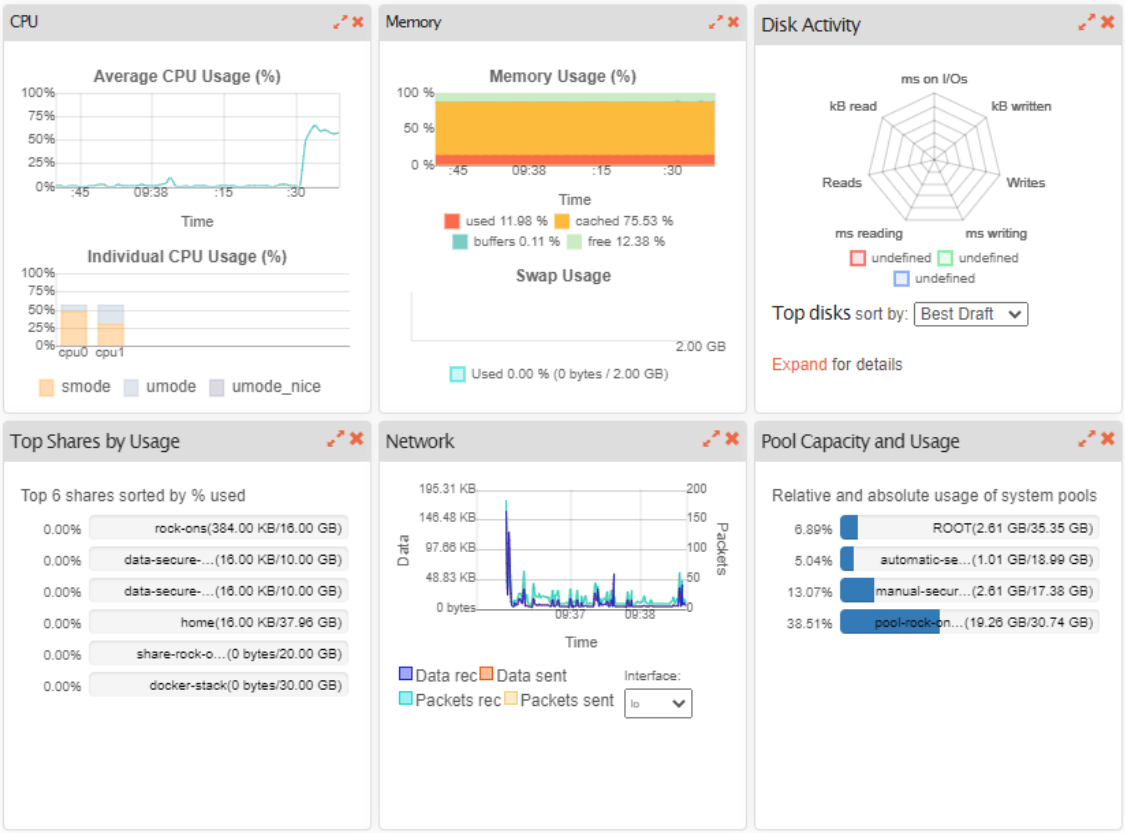
reálnom nasadení ostatných komponentov riešenia v prostredí kontajnerov boli potreby pamäte RAM zaznamenané pomocou príkazu `docker stats -all`, a vypísané do tabuľky 9.3.

Tab. 9.3: Približné reálne využitie pamäte RAM.

	Využitie RAM
Systém Rockstor	800 MB
Portainer	20 MB
Nextcloud	150 MB
MySQL	350 MB
ClamAV	550 MB
Archivematica Redis	5 MB
Archivematica Gearman	5 MB
Archivematica Fits	75 MB
Archivematica Elasticsearch	840 MB
Archivematica Nginx	5 MB
Archivematica MCP Server	55 MB
Archivematica MCP Client	65 MB
Archivematica Dashboard	250 MB
Archivematica Storage Service	80 MB
SignServer	700 MB
SignServer MariaDB	60 MB
EJBCA	850 MB
EJBCA MariaDB	70 MB
Spolu	**4930 MB**

V nasadení prostredníctvom virtuálneho stroja pridelené hardvérové zdroje postačovali na prácu so všetkými komponentami. Pri pomalších zariadeniach, alebo v prostredí s viac používateľmi, je už na zváženie využitie komponentov EJBCA a SignServer, ktoré majú dosť veľké systémové požiadavky. Spracovania údajov systémami Nextcloud a Archivematica závisí aj od rýchlosti použitého média, na ktoré sa dané dáta ukladajú.

Monitoring jednotlivých kontajnerov sa dá vykonať aj prostredníctvom Portainera, kde sú všetky informácie prehľadne graficky zobrazené. Pre monitoring celého systému má Rockstor na hlavnej stránke webového rozhrania (Dashboard) implementované grafické zobrazenia jednotlivých hlavných komponentov, viď obr. 9.18.



Obr. 9.18: Monitoring systémových zdrojov v Rockstor.

9.12 Zhodnotenie bezpečnosti dát

Základom realizácie návrhu je systém Rockstor. Ten využíva súborový systém Btrfs, ktorý je zameraný na implementáciu pokročilých funkcií, odolnosti voči chybám a ich ľahkú opravu. Taktiež podporuje softvérový RAID, ktorý je využitý na zrkadlenie celých diskov (RAID 1) pre zamedzenie straty dát pri poruche jedného z diskov. Samotný systém počíta s využitím až 6 samostatných diskov, kde sú vytvorené dve zrkadlené dvojice. Prvá dvojica je určená pre bežné dáta a je vybavená LUKS 1 s automatickým odomykaním pri štarte systému. Druhá dvojica je určená pre citlivé dáta s LUKS 2 v manuálnom režime. To znamená, že k dátam bude prístup len vtedy keď to bude niekto požadovať a správca manuálne sprístupní dané úložisko. Po vykonaní všetkých potrebných úkonov ho zase ihneď uzamkne.

Samotný archivačný proces v prostredí Archivematica kontroluje všetky dáta antivírusovým systémom ClamAV. Vykonáva kontrolné súčty a pripravuje dáta podľa modelu OAIS a štandardu balíkov BagIt pre dlhodobé uchovanie. Vytvára tiež množstvo logov a metadát, ktoré sú pribalené k archivovaným dátam, aby bola zabezpečená ich spätná transformácia aj v budúcnosti. Práve podrobné údaje o archívnom procese spolu s kontrolnými súčtami a metadátach o autorovi sú zhrnuté

v balíku AIP v súbore METS podľa štandardu PREMIS. Ten slúži na kontrolu autenticity balíka a jeho obsahu a preto sa jeho kópia odporúča uložiť na iné separátne externé médium. Kvôli zvýšeniu dôveryhodnosti daných metadát je dobré obstarať tento XML súbor digitálnym podpisom s časovou pečiatkou XAdES-T.

V systéme Nextcloud, ktorý je hlavnou vstupnou bránou pre všetkých používateľov, je striktná politika hesiel, s nastavením minimálnej dĺžky (8 znakov) a typu znakov, expirácie hesla (60 dní) a počet pokusov o prihlásenie pred zablokovaním účtu (5). Taktiež sa nahrávané súbory kontrolujú antivírusovým systémom ClamAV. Bezpečnosť dopĺňajú prehľadné logy a možnosti monitoringu aktivít v celom systéme z pohľadu správcu.

9.12.1 Kryptografická bezpečnosť riešenia

Celé riešenie je navrhnuté a realizované tak, že vyžaduje využitie len publikovaných a overených kryptografických algoritmov a protokolov, ktoré sú považované v terajšej dobe za bezpečné. Použitá dĺžka kľúčov by mala byť minimálne podľa NIST a ECRYPT pre aktuálne obdobie. Pri LUKS 1 je použitá šifra AES v móde XTS s veľkosťou kľúča 256 bitov s odvodením kľúča od hesla pomocou PBKDF2 (SHA-256). Pre citlivé dáta bola zvolená vyššia úroveň bezpečnosti pomocou LUKS 2, ktorá používa šifru AES v móde XTS, veľkosť kľúča 512 bitov s odvodením kľúča od hesla pomocou Argon2i. Pri archivovaných balíkoch je ale nutné využitie čo najsilnejšej možnej varianty z dostupných kryptografických algoritmov, hlavne pri uchovávaní dát na dlhú dobu. Preto je v systéme Archivematica vybraný algoritmus pre kontrolné súčty SHA-512.

Výsledné riešenie sa považuje za bezpečné. Spĺňa požadované štandardy a odporúčania pri archivačnom procese, ako aj pri použití kryptografických algoritmov a ich dĺžku kľúčov. Netreba zabúdať, že je dôležitá aj kvalitná správa riešenia, vírusová kontrola, dodržiavanie pokynov používateľmi zo strany administrátorov, ako aj iné skutočnosti, ktoré dotvárajú bezpečnosť riešenia ako celok.

Záver

Jedným z cieľov tejto práce bola analýza možností realizácie privátneho cloudu a zabezpečenej archivácie dát pomocou riešení s otvoreným zdrojovým kódom. Táto analýza bola rozdelená na dve časti, kde prvá sa venuje výhradne privátnym cloudom. Z jednotlivých popisovaných a porovnávaných kritérií existujúcich vybraných riešení sa javí ako vhodné použitie systému Nextcloud. Spĺňa viaceré predpoklady na vytvorenie zabezpečeného, dobre spravovateľného privátneho cloudu na vlastnom zariadení. Druhá časť sa venuje archivačným systémom, kde vhodné riešenie by malo odzrkadľovať štandard OAIS (pre definovanie základných procesov dlhodobej archivácie), ako aj rôzne formáty metadát (METS, PREMIS, Dublin Core) pre zabezpečenie čitateľnosti. Tieto skutočnosti najlepšie spĺňa systém Archivematica. Z oboch častí analýzy tiež vyplynulo, že dôležitá je aktivita zo strany vývojárov, ako aj existencia širokej komunitnej základne či podrobnej dokumentácie. Rozšírenie základnej funkcionality, či dobrá možnosť kustomizácie pre potreby používateľov, je tiež vítaná vlastnosť.

Druhým cieľom práce bolo navrhnúť privátny cloud s podporou dlhodobej archivácie. Ako základ pre implementáciu bola navrhnutá odľahčená virtualizácia prostredníctvom izolácií jednotlivých systémov do kontajnerov. Týmto spôsobom sa očakáva nasadenie privátneho cloudu po boku archivačného systému, na základe výsledkov počiatočným analýz. V návrhu je popísaný celý archivačný proces, ktorý sa zakladá na modeli OAIS. Súčasťou archivačného procesu by malo byť vytvorenie kontrolných súčtov pre zabezpečenie integrity ako aj generovanie komplexných metadát METS a PREMIS s možnosťou ich extrakcie. Z hľadiska bezpečnosti je potrebné, aby riešenie využívalo iba bezpečné kryptografické algoritmy a protokoly.

Na základe návrhu boli sformované štyri koncepty implementácie, kde boli otestované jednotlivé archivačné systémy v nasadení s privátnym cloudom Nextcloud. Výstupom je porovnanie a ohodnotenie jednotlivých konceptov na základe kritérií, ktoré odrážajú oblasti funkcionality, obsluhy a správy. Najviac bodov získal koncept číslo 4 s archivačným systémom Archivematica.

Posledným hlavým bodom bola implementácia riešenia na základe analýz, návrhu a testovania. Jadro riešenia tvorí systém Rockstor, ktorý sa stará o dôvernosť dát pomocou šifrovania celých diskov LUKS. Taktiež je použité zrkadlenie RAID 1 pre ochranu proti náhlej strate údajov. Na Rockstore bežia ostatné systémy vo forme kontajnerov. Ich príprava a nasadenie je podrobne popísaná. Výsledok tvorí sústava 16 kontajnerov, ktoré sú nasadené spoločne ako stack. Základ tvorí archivačný systém Archivematica s privátnym cloudom Nextcloud. Archivácia prebieha spôsobom, že používateľ Nextcloudu vyberie dáta pre uchovanie a presunie ich na vyhradené miesto. Následne v Archivematike potvrdí výber dát a zvolí ako ich archivovať. Celý

proces prebieha automaticky, až po uloženie výsledného AIP balíka do zabezpečeného úložiska pomocou LUKS. Všetky vykonané akcie sa ukladajú s časovými záznamami vo forme metadát do XML súboru METS. Pre možnosti podpisovania dokumentov je nasadený podpisový server SignServer, ktorý dokáže digitálne podpísať napríklad vyextrahovaný súbor metadát z balíka AIP, METS štandardom XAdES-T.

Z pohľadu náročnosti implementácie na hardvér, bolo pri nasadení potreba 5 GB pamäte RAM. Pre plynulý chod sa vyžadujú aspoň dve jadrá procesora a 8 GB pamäte RAM. Hardvérové nároky závisia od počtu používateľov ako aj veľkosti dát, s ktorými sa pracuje. Z bezpečnostného hľadiska zaisťuje výsledné riešenie rôznymi spôsobmi a funkciami integritu, autentickosť, dôvernosť, nepopierateľnosť a časové ukotvenie dát. Všetky komponenty návrhu používajú iba publikované a overené kryptografické algoritmy a protokoly. Nesmú teda využívať prelomiteľné algoritmy, či nie moc bezpečné módy šifier. Pre túto skutočnosť je riešenie považované za kryptograficky bezpečné. Celková bezpečnosť sa odráža aj od dodržiavania určitých bezpečnostných zásad. Najväčšou hrozbou tak ostáva používateľ, ktorý často nedodržiava bezpečnostné pokyny.

Literatúra

[1] KIRSCH, Daniel a Judith HURWITZ, 2020. *Cloud Computing For Dummies.* 2nd Edition. New Jersey: John Wiley. ISBN 978-1-119-54665-8.

[2] NEWCOMBE, Lee, 2020. *Securing Cloud Services: A pragmatic guide. Second edition.* United Kingdom: IT Governance Publishing. ISBN 978-1-78778-207-5.

[3] MALISOW, Ben, 2020. *(ISC)2 CCSP Certified Cloud Security Professional: Official Study Guide.* USA: John Wiley. ISBN 978-1-119-60337-5.

[4] CHEN, Lei, Hassan TAKABI a Nhien-an LE-KHAC, 2019. *Security, Privacy, and Digital Forensics in the Cloud.* Singapore: John Wiley. ISBN 9781119053408.

[5] D'AGOSTINO, Giulio, 2019. *Data Security in Cloud Computing.* Volume I. New York: Momentum Press. ISBN 978-1-94708-399-8.

[6] LE, Dac-Nhuong, Raghvendra KUMAR, Gia Nhu NGUYEN a Jyotir Moy CHATTERJEE, 2018. *Cloud computing and virtualization.* USA: John Wiley. ISBN 978-1-119-48790-6.

[7] MCKENDRICK, Russ, 2020. *Mastering Docker: Enhance your containerization and DevOps skills to deliver production-ready applications.* Fourth edition. UK: Packt Publishing. ISBN 978-1-83921-657-2.

[8] BUGNION, Edouard, Jason NIEH a Dan TSAFRIR, 2017. *Hardware and software support for virtualization.* SYNTHESIS LECTURES ON COMPUTER ARCHITECTURE, Lecture 38. Morgan & Claypool. ISBN 9781627056939.

[9] SCHENKER, Gabriel N., Hideto SAITO, Hui-Chuan Chloe LEE a Ke-Jou Carol HSU, 2019. *Getting Started with Containerization: Reduce the Operational Burden on Your System by Automating and Managing Your Containers.* UK: Packt Publishing. ISBN 978-1-83864-570-0.

[10] *What is Docker?* [online]. [cit. 2021-5-23]. Dostupné z: <`https://www.ibm.com/cloud/learn/docker`>

[11] *Docker Documentation* [online]. [cit. 2021-5-23]. Dostupné z: <`https://docs.docker.com/`>

[12] *Portainer Open Source Container Management GUI for Kubernetes, Docker, Swarm* [online]. [cit. 2021-5-23]. Dostupné z: <`https://www.portainer.io/`>

[13] BAUER, Roderick. *What's the Diff: Backup vs Archive.* Backblaze [online]. August 2, 2018 [cit. 2020-12-11]. Dostupné z: <https://www.backblaze.com/blog/data-backup-vs-archive/>

[14] *BACKUP VS ARCHIVE: WHAT'S THE DIFFERENCE AND WHY YOU NEED BOTH* [online]. USA: Waters Corporation, November 2015 [cit. 2020-12-11]. Dostupné z: <https://www.scientific-computing.com/sites/default/files/NG_Backup_vs_Archive_WP_0.pdf>

[15] *Archive Disaster Recovery* [online]. UK: Micro Focus, 2017 [cit. 2020-12-11]. Dostupné z: <https://www.microfocus.com/media/data-sheet/archive_disaster_recovery_ds.pdf>

[16] *Levels of Digital Preservation* [online]. The National Digital Stewardship Alliance, 2018 [cit. 2020-12-11]. Dostupné z: <https://ndsa.org/publications/levels-of-digital-preservation/>

[17] LAVOIE, Brian. *The Open Archival Information System (OAIS) Reference Model: Introductory Guide* [online]. Digital Preservation Coalition, October 2014, (Second edition) [cit. 2020-12-11]. ISSN 2048-7916. Dostupné z: <http://dx.doi.org/10.7207/twr14-02>

[18] *Recommendation for Space Data System Practices: REFERENCE MODEL FOR AN OPEN ARCHIVAL INFORMATION SYSTEM (OAIS)* [online]. Washington, DC, USA: Consultative Committee for Space Data Systems, June 2012 [cit. 2020-12-11]. Dostupné z: <https://public.ccsds.org/pubs/650x0m2.pdf>

[19] *Pre-Ingest Tool* [online]. [cit. 2021-5-23]. Dostupné z: <https://coptr.digipres.org/index.php/Pre-Ingest_Tool>

[20] BACA, Murtha, 2016. *Introduction to metadata* [online]. Third edition. Los Angeles: Getty Research Institute [cit. 2020-12-11]. ISBN 9781606064795. Dostupné z: <http://www.getty.edu/publications/intrometadata/>

[21] HAYNES, David, 2018. *Metadata for Information Management and Retrieval: Understanding metadata and its use.* Second edition. London: Facet Publishing. ISBN 978-1-85604-824-8.

[22] RILEY, Jenn, 2017. *UNDERSTANDING METADATA: WHAT IS METADATA, AND WHAT IS IT FOR?.* Baltimore, MD: National Information Standards Organization. ISBN 978-1-937522-72-8.

[23] *PREMIS Data Dictionary for Preservation Metadata* [online], 2015. Version 3.0. PREMIS Editorial Committee [cit. 2020-12-11]. Dostupné z: <https://www.loc.gov/standards/premis/>

[24] DAPPERT, Angela, Sébastien PEYRARD a Rebecca Squire GUENTHER, ed., 2016. *Digital Preservation Metadata for Practitioners: Implementing PREMIS*. Switzerland: Springer International Publishing. ISBN 978-3-319-43761-3.

[25] *Dublin Core Metadata Initiative: innovation in metadata design, implementation & best practice* [online]. [cit. 2020-12-11]. Dostupné z: <https://www.dublincore.org/specifications/dublin-core/>

[26] *Metadata Encoding & Transmission Standard* [online]. [cit. 2020-12-11]. Dostupné z: <https://www.loc.gov/standards/mets/>

[27] *Encoded Archival Description Tag Library: Version EAD3* [online], 2015. Chicago, USA: Society of American Archivists [cit. 2021-5-23]. ISBN 1-931666-89-X. Dostupné z: <https://www2.archivists.org/sites/all/files/TagLibrary-VersionEAD3.pdf>

[28] *BagIt* [online]. [cit. 2021-5-23]. Dostupné z: <http://fileformats.archiveteam.org/wiki/BagIt>

[29] *The BagIt File Packaging Format (V1.0)* [online]. In: . Library of Congress, October 2018 [cit. 2021-5-23]. ISSN 2070-1721. Dostupné z: <https://datatracker.ietf.org/doc/html/rfc8493>

[30] *E-ARK Project* [online]. [cit. 2021-5-23]. Dostupné z: <https://www.eark-project.com/>

[31] *E-ARK4ALL Project* [online]. [cit. 2021-5-23]. Dostupné z: <https://e-ark4all.eu/>

[32] *Common Specification for Information Packages* [online]. [cit. 2021-5-23]. Dostupné z: <https://dilcis.eu/specifications/common-specification>

[33] *E-ARK SIP: Specification for Submission Information Packages* [online]. 12.06.2020 [cit. 2021-5-23]. Dostupné z: <https://earksip.dilcis.eu/pdf/eark-sip.pdf>

[34] NOTENBOOM, Leo A. *What's the Best Long-Term Storage Media?: Tips to avoid losing data in your lifetime* [online]. [cit. 2021-5-23]. Dostupné z: <https://askleo.com/best-long-term-storage-media/>

[35] HENRIKSEN, Sofie Laier, Wiel SEUSKENS a Gaby WIJERS. *D6.2 Best practices for a digital storage infrastructure for the long-term preservation of digital files* [online]. Digitising Contemporary Art [cit. 2021-5-23]. Dostupné z: <https://pro.europeana.eu/files/Europeana_Professional/Projects/Project_list/Digitising_Contemporary_Art/Deliverables/DCA_D62_Best_practices_for_a_digital_storage_infrastructure_20130506_Version1.pdf>

[36] *Disk Storage vs. Tape Storage* [online]. [cit. 2021-5-23]. Dostupné z: <http://www.tape-storage.net/en/storage_comparison/article_01/>

[37] *Btrfs* [online]. [cit. 2021-5-23]. Dostupné z: <https://wiki.debian.org/Btrfs>

[38] *Btrfs Wiki* [online]. [cit. 2021-5-23]. Dostupné z: <https://btrfs.wiki.kernel.org/index.php/Main_Page>

[39] BAROŇÁK, IVAN. *RAID: ČO JE TO A AKO NÁM POMÁHA NEPRÍSŤ O DÁTA* [online]. 18. MARCA 2019 [cit. 2021-5-23]. Dostupné z: <https://netvel.sk/co-je-to-raid/>

[40] *Dm-crypt/Device encryption* [online]. [cit. 2021-5-23]. Dostupné z: <https://wiki.archlinux.org/title/Dm-crypt/Device_encryption>

[41] BROŽ, Milan. *LUKS2: On-Disk Format Specification* [online]. 23.10.2018 [cit. 2021-5-23]. Dostupné z: <https://gitlab.com/cryptsetup/cryptsetup/blob/master/docs/on-disk-format-luks2.pdf>

[42] *Announcing the ADVANCED ENCRYPTION STANDARD (AES)* [online], 2001. National Institute of Standards and Technology [cit. 2020-12-11]. FIPS PUB 197. Dostupné z: <https://nvlpubs.nist.gov/nistpubs/fips/nist.fips.197.pdf>

[43] DAEMEN, Joan a Vincent RIJMEN, 2002. *The Design of Rijndael: AES - The Advanced Encryption Standard.* 1. Springer. ISBN 3540425802.

[44] SMART, Nigel P., ed. *ECRYPT – Coordination & Support Action: Algorithms, Key Size and Protocols Report (2018).* H2020-ICT-2014 – Project 645421 [online]. 28 February 2018 [cit. 2020-12-11]. Dostupné z: <https://www.ecrypt.eu.org/csa/documents/D5.4-FinalAlgKeySizeProt.pdf>

[45] *XTS Mode* [online]. [cit. 2021-5-23]. Dostupné z: <https://www.cryptopp.com/wiki/XTS_Mode>

[46] ROTT, Jeffrey Keith. *Intel Advanced Encryption Standard Instructions (AES-NI)* [online]. 02.03.2012 [cit. 2021-5-23]. Dostupné z: <https://software.intel.com/content/www/us/en/develop/articles/intel-advanced-encryption-standard-instructions-aes-ni.html>

[47] *FEDERAL INFORMATION PROCESSING STANDARDS PUBLICATION: Secure Hash Standard (SHS).* FIPS PUB 180-4 [online]. Gaithersburg, MD: National Institute of Standards and Technology, August 2015 [cit. 2020-12-11]. Dostupné z: <http://dx.doi.org/10.6028/NIST.FIPS.180-4>

[48] Introduction to Digital Preservation: Fixity. *Bodleian Libraries: Oxford LibGuides* [online]. [cit. 2021-5-23]. Dostupné z: <https://libguides.bodleian.ox.ac.uk/digitalpreservation/fixity>

[49] STEFANO, Paula De, Carl FLEISCHHAUER, Andrea GOETHALS, Michael KJÖRLING a Nick KRABBENHOEFT. *What is Fixity, and When Should I be Checking It?* [online]. In: . 2014 [cit. 2021-5-23]. Dostupné z: <https://www.digitalpreservation.gov/documents/NDSA-Fixity-Guidance-Report-final100214.pdf>

[50] *Recommendation for Key Management: Part 1 – General.* NIST Special Publication 800-57 Part 1: Revision 5 [online]. National Institute of Standards and Technology, May 2020 [cit. 2020-12-11]. Dostupné z: <https://doi.org/10.6028/NIST.SP.800-57pt1r5>

[51] KATZ, Jonathan, 2010. *Digital Signatures.* USA: Springer Science. ISBN 978-0-387-27711-0.

[52] *FEDERAL INFORMATION PROCESSING STANDARDS PUBLICATION: Digital Signature Standard (DSS).* FIPS PUB 186-4 [online]. Gaithersburg, MD: National Institute of Standards and Technology, July 2013 [cit. 2020-12-11]. Dostupné z: <http://dx.doi.org/10.6028/NIST.FIPS.186-4>

[53] *eSignature Documentation: What is an electronic signature?* [online]. [cit. 2021-5-23]. Dostupné z: <https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/Introduction+to+e-signature>

[54] *Electronic Signatures and Infrastructures (ESI): XML Advanced Electronic Signatures (XAdES)* [online]. 2010-12 [cit. 2021-5-23]. Dostupné z: <https://www.etsi.org/deliver/etsi_ts/101900_101999/101903/01.04.02_60/ts_101903v010402p.pdf>

[55] *Electronic Signatures and Infrastructures (ESI): CMS Advanced Electronic Signatures (CAdES)* [online]. 2013-04 [cit. 2021-5-23]. Dostupné z: <https://www.etsi.org/deliver/etsi_ts/101700_101799/101733/02.02.01_60/ts_101733v020201p.pdf>

[56] *Electronic Signatures and Infrastructures (ESI): PDF Advanced Electronic Signature Profiles* [online]. 2009-07 [cit. 2021-5-23]. Dostupné z: <https://www.etsi.org/deliver/etsi_ts/102700_102799/10277801/01.01.01_60/ts_10277801v010101p.pdf>

[57] BRZICA Hrvoje, Boris Herceg, Stančić Hrvoje , November 2013. *Long-term Preservation of Validity of Electronically Signed Records.* Zagreb, Croatia: IN-Future2013

[58] *SignServer Product Documentation: Install, set up, and use SignServer* [online]. [cit. 2021-5-24]. Dostupné z: <https://www.signserver.org/documentation/>

[59] *Nextcloud: The self-hosted productivity platform that keeps you in control* [online]. [cit. 2020-12-11]. Dostupné z: <https://nextcloud.com/>

[60] *Security and authentication: White paper* [online]. Stuttgart Germany: Nextcloud, 2018 [cit. 2020-12-11]. Dostupné z: <https://nextcloud.com/secure/>

[61] *Owncloud: Store. Share. Work.* [online]. [cit. 2020-12-11]. Dostupné z: <https://owncloud.com/>

[62] *Data Protection and Data Secrecy in ownCloud: How Encryption can help!* WHITEPAPER [online]. Nürnberg Germany: ownCloud [cit. 2020-12-11]. Dostupné z: <https://owncloud.com/>

[63] *Seafile* [online]. [cit. 2020-12-11]. Dostupné z: <https://www.seafile.com/>

[64] *Seafile admin manual* [online]. [cit. 2020-12-11]. Dostupné z: <https://manual.seafile.com/>

[65] *Pydio: On Premises File Sharing Done Right* [online]. [cit. 2020-12-11]. Dostupné z: <https://pydio.com/>

[66] *Administration Guide for Cells v2* [online]. [cit. 2020-12-11]. Dostupné z: <https://pydio.com/en/docs/administration-guides>

[67] *Archivematica documentation: 1.8.1* [online]. [cit. 2020-12-11]. Dostupné z: <https://www.archivematica.org/en/docs/archivematica-1.8/>

[68] *Archivematica* [online]. [cit. 2020-12-11]. Dostupné z: <https://coptr.digipres.org/Archivematica>

[69] *Florida Digital Archive* [online]. [cit. 2020-12-11]. Dostupné z: <https://libraries.flvc.org/florida-digital-archive>

[70] CAPLAN, Priscilla a Carol C.H. CHOU. *DAITSS grows up: Migrating to a second generation preservation system* [online]. Gainesville, FL / USA: Florida Center for Library Automation, January 2011 [cit. 2020-12-11]. Dostupné z: <https://www.researchgate.net/publication/291282530_DAITSS_grows_up_Migrating_to_a_second_generation_preservation_system>

[71] *DAITSS (Dark Archive in the Sunshine State) software now available* [online]. [cit. 2020-12-11]. Dostupné z: <https://www.loc.gov/standards/premis/DAITSS-announcement.html>

[72] *RODA 3: LONG-TERM DIGITAL PRESERVATION.* CHARACTERISTICS AND TECHNICAL REQUIREMENTS: WHITE PAPER [online]. Keep, 29.11.2018 [cit. 2020-12-11]. Dostupné z: <https://www.keep.pt/wp-content/uploads/2019/12/WP181216-whitepaper-roda-3_EN.pdf>

[73] *Roda: Documentation* [online]. [cit. 2020-12-11]. Dostupné z: <https://github.com/keeps/roda/blob/master/documentation/README.md>

[74] *ES Solutions: ESSArch* [online]. [cit. 2021-5-24]. Dostupné z: <https://www.essolutions.se/essarch/>

[75] *ESSArch* [online]. [cit. 2021-5-24]. Dostupné z: <https://docs.essarch.org/index.html>

[76] SCHMIDT, Rainer, Sven SCHLARB a Björn SKOG. *E-ARK: European Archival Records and Knowledge Preservation* [online]. [cit. 2021-5-24]. Dostupné z: <http://www.eark-project.com/resources/project-deliverables/54-d62intplatformref-1/EARK_D6_2.pdf>

[77] *Rockstor* [online]. [cit. 2020-12-11]. Dostupné z: <http://rockstor.com/>

[78] *Documentation* [online]. Rockstor [cit. 2020-12-11]. Dostupné z: <http://rockstor.com/docs/index.html>

[79] *ESSolutions / ESSArch* [online]. [cit. 2021-5-24]. Dostupné z: <https://github.com/ESSolutions/ESSArch>

[80] *E-ARK-Software / earkweb* [online]. [cit. 2021-5-24]. Dostupné z: <https://github.com/E-ARK-Software/earkweb>

[81] *Artefactual-labs / am* [online]. [cit. 2021-5-24]. Dostupné z: <https://github.com/artefactual-labs/am/tree/master/compose#docker-and-linux>

Zoznam symbolov, veličín a skratiek

IT	Informačné technológie – Information technology
SaaS	Softvér ako služba – Software as a Service
PaaS	Platforma ako služba – Platform as a Service
IaaS	Infraštrukúra ako služba – Infrastructure as a Service
API	Aplikačné programové rozhranie – Application programming interface
LXC	Linux kontajnery – Linux Containers
YAML	YAML nie je značkovací jazyk – YAML Ain't Markup Language
ISO	Medzinárodná organizácia pre štandardy – International Organization for Standardization
NDSA	Národná aliancia pre digitálne správcovstvo – National Digital Stewardship Alliance
LoP	Úrovne digitálneho uchovávania – Levels of Digital Preservation
OAIS	Otvorený archívny informačný systém – Open Archival Information System
CCSDS	Poradný výbor pre vesmírne dátové systémy – Consultative Committee for Space Data Systems
PDI	Informácie o uchovaní – Preservation Description Information
SIP	Balík informácií na predloženie – Submission Information Package
AIP	Balík archívnych informácií – Archival Information Package
DIP	Balík informácií pre šírenie – Dissemination Information Package
IP	Balík informácií – Information Package
NISO	Národná organizácia pre štandardizáciu informácií – National Information Standards Organization
PREMIS	Metadáta uchovania: Stratégie implementácie – Preservation Metadata: Implementation Strategies
DC	Dublinské jadro – Dublin core

METS	Štandard kódovania a prenosu metadát – Metadata Encoding and Transmission Standard
XML	Rozšíriteľný značkovací jazyk – Extensible Markup Language
EAD	Zakódovaný archívny popis – Encoded Archival Description
TAR	Archívna páska – Tape ArchiveArchívna páska
MD5	Algoritmus Message-Digest 5 – Message-Digest algorithm 5
SHA	Zabezpečený algoritmus hash – Secure Hash Algorithm
E-ARK	Európske archivačné záznamy a znalosti uchovávania – European Archival Records and Knowledge Preservation
RODA	Úložisko autentických digitálnych záznamov – Repository of Authentic Digital Records
CSIP	Spoločná špecifikácia informačných balíkov – Common Specification for Information Packages
ID	Identifikačné číslo – Identification number
Btrfs	Súborový systém B-strom – B-tree file system
HDD	Pevný disk – Hard Disk Drive
SSD	Mechanika s nepohyblivým médiom – Solid state drive
CD	Kompaktný disk – Compact Disk
DVD	Digital Video Disc
Blu-ray	Modrý lúč – Blue Ray
DAT	Digitálna zvuková páska – Digital Audio Tape
WORM	Write Once Read Many
LTO	Jedno-nábojová páska – Linear Tape Open
CoW	Kópia pri zápise – Copy on Write)
GPL	Všeobecná verejná licencia – General Public License
RAID	Redundantné pole nezávislých diskov – Redundant Array of Independent Disks

Ext4	Štvrtý rozšírený súborový systém – Fourth extended filesystemExt4
LUKS	Zjednotené nastavenie systému Linux – Linux Unified Key Setup
LTO	Intel AES inštrukcie novej generácie – Intel AES New Instructions
PBKDF2	Funkcia odvodenia kľúča na základe hesla 2 – Password-Based Key Derivation Function 2
NTFS	Súborový Systém Novej Technológie – New Technology File System
ZFS	Z súborový systém – Z File System
AES	Pokročilý šifrovací štandard – Advanced Encryption Standard
ECB	Elektronická kódova kniha – Electronic Code Book
CBC	Reťazové šifrovanie blokov – Cipher block chaining
OFB	Výstupná spätná väzba – Output FeedBack
CTR	Režim počítadla – Counter Mode
XTS	Utrhnutý kódový zoznam založený na XEX – XEX-based tweaked-codebook
XEX	XOR šifrovanie XOR – XOR Encrypt XOR
XOR	eXclusive OR
CCM	Režim počítadla šifrovaných reťazených blokov MAC – Counter Mode Cipher Block Chaining MAC
GCM	Galois/Režim počítadla – Galois/Counter Mode
CWC	Carter–Wegman CTR mode
OCB	Režim kompenzácie kódov – Offset Codebook Mode
MAC	Overovací tag správy – Message authentication tag
DSA	Digitálny podpisový algoritmus – Digital Signature Algorithm
NIST	Národný inštitút pre štandardy a technológie – National Institute of Standards and Technology
ECRYPT	Európska sieť excelentnosti pre kryptológiu – European Network of Excellence for Cryptology

AdES	Zdokonalené elektronické podpisy – Advanced Electronic Signatures
XAdES	XML Zdokonalené elektronické podpisy – XML Advanced Electronic Signature
CMS	CMS Zdokonalené elektronické podpisy – CMS Advanced Electronic Signatures
PDF	PDF Zdokonalené elektronické podpisy – PDF Advanced Electronic Signatures
MRTD	Strojovo čitateľný cestovný doklad – Machine-readable travel document
ICAO	Medzinárodná organizácia pre civilné letectvo – International Civil Aviation Organization
ETSI	Európsky inštitút pre telekomunikačné normy – European Telecommunications Standards Institute
PDF	Prenosný formát dokumentu – Portable Document Format
SSL	Vrstva bezpečných soketov – Secure Sockets Layer
TLS	Zabezpečenie transportnej vrstvy – Transport Layer Security
AD	Aktívny adresár – Active Directory
LDAP	Ľahký adresárový prístupový protokol – Lightweight Directory Access Protocol
HTTPS	Zabezpečený hypertextový prenosový protokol – Hypertext Transfer Protocol Secure
NAS	Sieťové úložisko – Network Attached Storage
USB	Univerzálna sériová zbernica – Universal Serial Bus
HSM	Hardvérový bezpečnostný modul – Hardware security module
DAITSS	Temný archív v slnečnom štáte – Dark Archive in the Sunshine State
API	Rozhranie pre programovanie aplikácií – Application programming interface
MD4	Algoritmus Message-Digest 4 – Message-Digest algorithm 4

DES	Štandard šifrovania údajov – Data Encryption Standard
RC4	Rivest Cipher 4
MITM	Človek uprostred – Man in the middle
SSH	Zabezpečený príkazový interpretátor – Secure Shell
SCP	Zabezpečené kopírovanie – Secure Copy
LZMA	Lempel–Ziv–Markov chain algorithm
JAR	Archív Java – Java Archive
CPU	Centrálna procesorová jednotka – Central processing unit
RAM	Pamäť s náhodným prístupom – Random Access Memory
TB	Terabajt – Terabyte
GB	Gigabajt – Gigabyte
MB	Megabajt – Megabyte
GHz	Gigahertz

A Obsah digitálnej prílohy

Digitálna príloha obsahuje YAML súbor `docker-compose.yml`, ktorý slúži ako predloha konfigurácie riešenia v prostredí kontajnerov Docker, a je zabalený do archívu `.zip`.